1.0

4.5
5.0
5.6

2.8

2.5

3.2

2.2

3.6

1.1

4.0

2.0

1.8

1.25

1.4

1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFIT/DS/EE/78-4

LEVEL

(1)

AD A0 66194

D D C
MAR 22 1979
RECEIVED
C

(6) NEW TECHNIQUES FOR TRACKING SEQUENCES

OF

DIGITIZED IMAGES.

DISSERTATION

(14) AFIT/DS/EE/78-4   Llewellyn S. Dougherty
Captain   USAF

(9) Doctoral thesis

(11) Nov 78   (12) 217 p.

Approved for public release; distribution unlimited

012 225

79 03 19 018

NEW TECHNIQUES FOR TRACKING SEQUENCES

OF

DIGITIZED IMAGES

DISSERTATION

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air Training Command

in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

by

Llewellyn S. Dougherty, B.S., M.S.
Captain   USAF

Approved for public release; distribution unlimited

AFIT/DS/EE/78-4

NEW TECHNIQUES FOR TRACKING SEQUENCES

OF

DIGITIZED IMAGES

by

Llewellyn S. Dougherty, B.S., M.S.
Captain   USAF

Approved:

Gary B. Lamont _____   27 November 1978
Chairman

Matthew Kabrisky _____   27 November 1978.

Stanley R. Robinson _____   27 November 1978

John Jones Jr. _____   27 November 1978

James J. Silverthorn _____   27 November 1978

Accepted:

J.S. Przemieniecki _____   27 November 1978
Dean, School of Engineering

Acknowledgments

I would like to acknowledge my indebtedness to the many individuals who have individually and collectively provided the guidance and encouragement, reflective criticism and stimulating discussion which have contributed so substantially to the research presented here. I would like to thank Professor Gary Lamont for his invaluable encouragement and help. His timely advice proved most helpful. Professor Kabrisky served most admirably as chief philosophizer and soothsayer, and stimulated much of the early insight into the problem. Dr. Robinson was simultaneously the devil's advocate and chief marketing agent for the results of the research. Dr. Griswold provided constant encouragement to press forward in the face of uncertainty.

As in any project where the end result depends on the quality of the data, the individuals who provide or collect the data deserve their own special mention. Captain Felix Morgan of the Air Force Weapons Laboratory provided the AIRPLANE data sequence, and Mike Novisky of the Air Force Avionics Laboratory was responsible for both TREES and CARS.

Finally, I must say a special thanks to my wife Stephanie and daughter K.C. Without their constant sacrifice, understanding, and encouragement, success would still be not but a dream.

Table of Contents

List of Figures

x

List of Tables

ABSTRACT

A model for a generalized image tracking system is presented. Characteristics of minimum norm similarity detectors are investigated. A first order local tangent plane model for digitized imagery is used to successfully predict properties of the auto- and cross-distance functions for real data. A matrix signal-to-noise ratio is shown to be the natural signal-to-noise ratio for the minimum norm detection problem and an approximation is derived and experimentally verified for an upper bound on the probability that a minimum norm detector makes a particular registration error. A non-linear two-dimensional filter is presented which shows a significant reduction in noise variance in low contrast regions of an image. An optimum weighted norm is derived which minimizes the probability of making a registration error, and an adaptive reference set selection algorithm is presented which maximizes the tracking signal-to-noise ratio. The adaptive reference set selection algorithm uses the histogram of gradient magnitudes and includes a new gradient estimator/classifier with a fixed probability of error. An adaptive Kalman filter is developed to update the reference image and the filter is shown to be stable in all areas of interest. An integrated sequential image tracking algorithm using the non-linear two-dimensional filter, adaptive reference set selection, and the adaptive Kalman filter for reference update is shown to perform significantly better than a conventional minimum norm tracking algorithm.

Chapter 1

INTRODUCTION

This investigation is concerned with developing new techniques for tracking images: specifically, sampled and digitized images. The major thrust of this study is that there are a number of ways to improve a tracking system's performance in the presence of noise. Potential improvements in system performance can be traded off for lower system cost or retained for higher speed, improved accuracy, and a larger class of trackable images.

1.1 Research Objectives and Approach

The primary goals of this research are to develop and evaluate techniques for tracking sequences of digitized images. The approach to be used in searching for improved tracking techniques will be to first develop a meaningful signal-to-noise ratio by examining the characteristics of a minimum norm similarity measure for digitized images, and then investigate ways to increase the signal-to-noise ratio through "intelligent" processing. By adopting a model for a generalized image-tracking system which is partitioned by task, and examining each task for ways to improve the effective signal-to-noise ratio, the overall system performance may be improved. It is essential to recognize the importance of interactions among the various tasks or operations within the tracking system and to avoid employing innovations in one area which might be detrimental to the total system.

1.2 The Image Tracking Task

Consider for a few moments the problems of building a machine to play tennis. Obviously, there are a number of functions that must be carried out by this machine if it is to do any more than simply serve the ball. It must certainly be able to move around the court with sufficient speed to return the ball; it must be able to swing a racket and hit the ball; but above all, it must be able to track the ball with sufficient accuracy to be able to predict where and when to swing the racket.

If we build this machine with an electronic eye, how will it actually track the ball? To answer this question, we must investigate the various pieces of the general machine-implemented image-tracking task.

The tracking task can be thought of as consisting of two separable and sequential subtasks: acquiring the target and tracking the acquired target image. Before an object or target can be tracked, it must be located within the image. For a human, this may be simple, but even this almost unconscious action consists of a sequence of tasks which are quite difficult for a machine. An orderly search must be conducted using stored data about where the target is likely to be. Then, for each object found, the recognition or classification task must be attempted, perhaps with the aid of contextual information. This problem, while it may be trivial for a 10 year old, is very difficult to solve using a machine. In fact, for most current applications, a human operator must still perform the acquisition or target designation task for the tracking machine.

1.3   A Model of an Image-Tracking System

Once an object to be tracked is found in the image, how is the actual tracking accomplished? The particular algorithm used by the human eye/brain combination is not known, but one possible model of a generalized image-tracking system is illustrated in Figure 1. The components of the model are discussed in the following paragraphs to provide some insight into the problems of mechanizing an image-tracking machine. The sequence of steps to be followed by the machine is as follows:

    I.   Reference image initialization

    II.   Image sensing

    III. Image motion compensation

    IV.   Preprocessing, consisting of

        A. Filtering

        B. Enhancement

        C. Feature extraction

        D. Comparison set selection

    V.   Similarity detection

    VI.   Reference image update

    VII. Target modeling and target state prediction

    VIII. Sensor movement to maintain the target in the field of
            view

In the following subsections, each of these steps will be discussed separately. Due to the interrelationships which exist between the various steps, some definitions will be deferred to each subsection.

3

PREPROCESSOR:

FILTERING
ENHANCEMENT
FEATURE EXTRACTION

IMAGE MOTION
COMPENSATION

SENSOR

INPUT IMAGE

SENSOR MOTION

RAW
DATA IMAGE

MOTION COMPENSATED
DATA IMAGE

MOTION COMPENSATED
PREPROCESSOR DATA IMAGE

SIMILARITY
DETECTION

REFERENCE IMAGE

INITIAL REFERENCE IMAGE

TARGET MODELING
AND
STATE PREDICTION

SENSOR DRIVE
COMMANDS

LOCATION OF MATCH

QUALITY OF MATCH

BEST MATCH
DATA IMAGE

REFERENCE
IMAGE
UPDATE

UPDATED REFERENCE IMAGE

Figure 1. Model of an Image Tracking System

4

1.3.1   Reference Image Initialization

The solution of the acquisition problem implies that an initial reference image exists since there must be something with which to compare the data image. For some systems this may be a photograph or a computer generated image of the target; it could be a prestored set of target characteristics; or, it might be one of the raw sensor images. In order to track the largest possible class of images, it seems essential that the initial reference be derived from the raw sensor data.   In Chapter 5 we will concentrate on extracting the best possible reference image from the sensor data.

The ability of a tracking system to maintain track on a particular target is strongly dependent on the quality of the reference image. If the initial reference image is not similar to the current sensed data image due to image motion or rapidly changing viewing angle, the similarity detection process (see Section 1.2.5) may not find a "best match" which is satisfactory, and the updated reference image may not represent the true target any more closely than did the initial reference image. In other words, a "trackable" target image is defined by having a reference image sequence which "converges" to the target image.

1.3.2   Image Sensing

Many different types of sensors can be used in image-tracking systems: vidicons, Correlatrons(1), charge-coupled devices (CCD's), charge-injection devices (CID's), one-dimensional and two-dimensional photodiode arrays, single-detector line-scanning systems, and many more

(1) A registered trademark of the Goodyear Corp.

5

that are still in the early stages of development. Among the sensors available today, there are a variety of scanning methods and sensitivities with a wide range of resolution, geometric fidelity, spectral response, and speed. The noise characteristics of the various types of sensors also vary between types. In some cases, the noise characteristics can vary within a single device due to environmental or operating conditions, or due to manufacturing methods. For example, a large area two-dimensional array sensor might be made up of many smaller arrays, each of which has a characteristic noise which is uniform over that small device; or a CCD sensor may have a nonuniform sensitivity over its active area, thus producing what is known as "fixed pattern noise" [7]. As a result, care must be taken in designing a tracking algorithm or preprocessor to assure compatibility with the sensor.

1.3.3   Image Motion Compensation

In the general case, the sensor may be both translating and rotating with respect to the target, but for the cases that we will be investigating, it is assumed that image motion on a frame-to-frame basis consists of translation only. There are a number of reasons why this assumption seems to be well founded for a large fraction of the tracking problems of interest. First, sensors that are stabilized in either pitch and yaw or azimuth and elevation on a moving vehicle will observe only small components of roll about the sensor optical axis as long as the vehicle is roll stabilized [21]. Second, when the sensor optical axis is pointed away from the instantaneous angular velocity vector by more than one half of the sensor instantaneous field of view,

6

the center of rotation for the image is not contained within the image itself. Thus, the resulting apparent image motion is in large part translation perpendicular to the vector from the center of the sensor field of view toward the instantaneous center of rotation of the image (see Figure 2). Third, for reasons that will be discussed later (see Section 2.2.1) the majority of the sensor instantaneous field of view is masked out by the tracker and only a small subsection of the available image is passed to the similarity detector. This small region is referred to as the field of regard or gated region of the image. This process results in a considerably smaller effective field of view than even the instantaneous field of view of the sensor, thus minimizing the error that can occur when correcting for rotation through translational adjustments.

An important part of any image-tracking scheme is the sensing of and compensation for the sensor motion. The size of the search region within an image depends on the uncertainty in the relative motion of the target with respect to the sensor optical axis. Any data which can be used to reduce the relative motion, or compensate for it, will have the effect of limiting the required search area for the similarity detection process. Reduction of the search area is directly translatable into an increase in the available computation time for each search point since fewer comparisons need to be made for each new data frame.

1.3.4 Preprocessing

The preprocessing step in an image-tracking system has historically been necessary because of the relatively poor performance of the similarity detection process when operating on raw imagery [8].

7

FRAME N+1

FRAME N

AVERAGE TRANSLATIONAL
COMPONENT OF MOTION

TRACKER FIELD
OF REGARD

SENSOR INSTANTANEOUS
FIELD OF VIEW AT
FRAME N

INSTANTANEOUS
CENTER OF
ROTATION
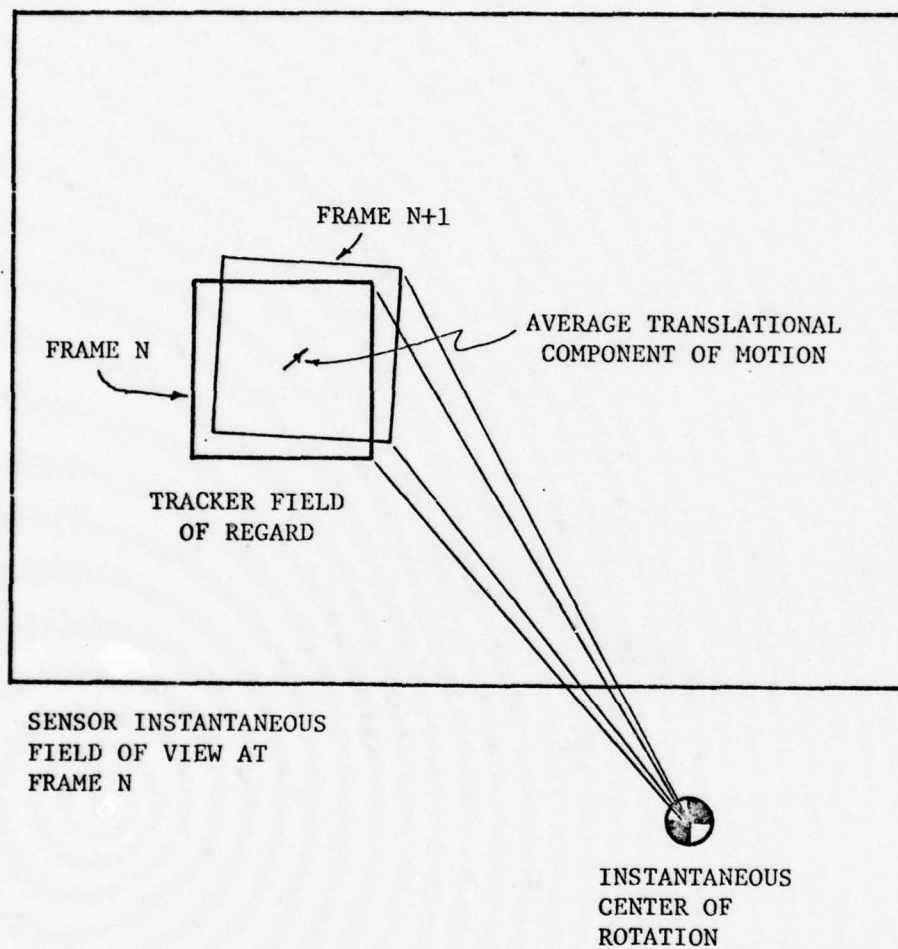
Figure 2.   Image Motion Due to Rotation About a Point
External to the Instantaneous Field of View

8

Typical preprocessing functions are amplitude normalization, signal
enhancement, noise reduction through linear and nonlinear filtering,
Walsh or Fourier transformation, and other feature extraction
techniques. The resulting image is generally tailored to the
requirements of the similarity detection process. The design of the
preprocessing function is still more of an art than a science, and the
success of most of the presently available image-tracking systems can
be attributed in large part to the insight of the preprocessor design
engineer into the characteristics of the sensor, the operating
environment, the typical imagery that would be encountered in the
field, and the similarity detection algorithm that is to be employed.
It is significant to note that the noise statistics in the raw data
image are generally determined by the sensor physics and the signal
detection and processing electronics. But, whatever the characteristics
of the sensor noise, the preprocessor will modify them. This fact also
enables the system designer to tailor the noise statistics if the
result is compatible with the remainder of the image-tracking process.
For example, noise in a sampled raw data image can be made
approximately Gaussian by processing the raw data image with a moving-
window averaging operator if the window contains a sufficient number of
points for the central limit theorem to be applied [26].

1.3.5   Similarity Detection

The similarity detection process compares the reference image with a
number of candidate subimages taken from the data image. The result of
each comparison is a similarity measure. The location of the subimage
which produces the maximum similarity measure (see Sections 2.2.1,

9

2.2.2, and 3.9) is taken as the current estimate of the location of the "target" as represented by the most recent reference image. The similarity detection process is the heart of any image-tracking algorithm. Some tracking algorithms produce only a direction from the center of the data image to the assumed location of the best match subimage without actually finding the best match (characteristic of most analog point-trackers, see Section 2.2.1); other tracking algorithms carry out the search for the most similar subimage and output its true location.

An exhaustive examination of the subimages within the search region requires a period of time which grows linearly with the area of the search region or as the square of the search radius. As a result, any new similarity detection algorithm which improves the speed for a single similarity measurement operation will allow either an increase in the rate at which data frames can be processed for a fixed search region size, or an increase in the size of the search region for a given data image rate.

1.3.6   Reference Image Update

A fixed reference image may be possible for those systems which are required to track objects which do not change their characteristics or which can be modeled as not changing their characteristics while they are being tracked. For our tennis-playing machine, it might be possible to model the tennis ball as a circle which never changes size and adjust the scale factor of the sensor output to make the received image of the ball match the reference image. An alternate approach is to change the reference image to incorporate the apparent change in the

10

size of the tennis ball. A refinement of the simple circular model for the tennis ball might include the location of the seams. This might make possible the estimation of the spin on the ball but would require that the reference image be updated to reflect the change in the location of the seams. The spin would then be estimated from the frame-to-frame change in the location and orientation of the seams. For an image-tracking system which maintains an actual image as the reference rather than a table of characteristics, the problem of updating the reference is equally complex. The basic decision that must be made is how much of an observed difference between the reference image and a data image is due to a change in the actual scene and how much is due to noise. In the past, the quality of the match between the reference image and the data image has been used to indicate when the reference image needed to be updated. The actual process of updating the reference has varied with the particular system. Some systems update by getting the next reference image from a storage file [37]. This type of system is obviously limited to tracking objects and images that are both definable ahead of time and not changing with time. Other systems, once they decide to update the reference, simply use the most recent data image as a new reference image [40]. In general, updating the reference image in a system without prestored reference data is a difficult problem. The criteria for evaluating the update scheme vary from system to system depending on the particular application. The update problem is one of measuring the changes between two images that are declared to be "most similar" by the similarity detector and deciding which changes are real and which changes are due to noise phenomena. The similarity detector compares the given reference with a

11

set of trial images and picks the "most similar" image for use by the reference update algorithm. For this reason, the reference image update task is treated as separable from, and independent of, the similarity detection task.

1.3.7   Target Modeling And Target State Prediction

An output of the similarity detection block in Figure 1 is the location of the "best match" between the reference image and the data image. In a classic feedback control system, this output signal would be used to drive sensor pointing angles directly. In a more modern "aided" tracking system, the target position signal is compared with a predicted target position, based on a model of both the sensor dynamics and the target dynamics. The difference is used to modify the state estimates rather than to drive the sensor directly. The control system then computes the appropriate drive signals to null the observed error based on the new updated state estimates and measured target position. If the model of the system (sensor and target) is perfect, there should be no observed error, and the models will not need to be updated. However, system noise and unmodeled states will produce errors which will disturb the models so that they are constantly in need of adjustment. Errors in the target model can lead to a requirement for an enlarged search area in the similarity detection process in order to compensate for bad (noisy) predictions of future target position. This increase in search area can result in a longer period of time being required to find the "best match" image. This increase in solution time then requires a longer prediction time with a further increase in the possible error, and thus, an even larger search area. Thus, computation

12

speed in the similarity detection process may define the maximum allowable search area, and any additional computation time taken by modeling must be compensated for through a reduced search area.

1.4 Summary of Report

In this chapter the research objectives have been outlined, the image tracking task has been defined, and the component parts of a generalized image-tracking system have been discussed. Chapter 2 presents as background material a summary of applications for image tracking systems, and some classical similarity detection techniques. In Chapter 3, a two-dimensional signal-to-noise ratio is developed for minimum norm similarity detectors and an upper bound is developed and experimentally verified for the probability of error associated with a particular type of misregistration. Chapter 4 presents three techniques for enhancing the signal-to-noise ratio, Chapter 5 presents an adaptive Kalman filter which is shown to be very helpful in reducing misregistration errors, and Chapter 6 develops and analyzes the performance of an integrated tracking algorithm which incorporates a complementary set of component algorithms. Chapter 7 provides a summary of this research and some suggestions for future work.

Chapter 2

BACKGROUND

In this chapter, some historical and current applications for image tracking systems are presented, and some classical similarity detection techniques are discussed. Much of the historical information presented in this chapter is drawn from the author's personal experiences over a ten year period in development, test, and evaluation of image tracking systems at the Air Force Missile Development Center and at the Air Force Avionics Laboratory.

2.1   Applications of Image-Tracking Systems

This section will briefly discuss some historical and current applications of image-tracking systems in an attempt to provide some insight and background for the unfamiliar reader, and perhaps relate this research to his experience. The reader is encouraged to think of some application for an image-tracking system which is peculiar to his field of expertise. For the purpose of enumeration here, image-tracking systems are divided into categories by the location of the tracking device:

14

I. Space Applications

    A. Automatic rendezvous with noncooperative targets

    B. Autonomous orbit determination

    C. Aiding earth based tracking for outer planet probes and orbiters

    D. Precision pointing for remote sensing systems

II. Airborne Applications

    A. Air-to-surface missile guidance

    B. Air-to-air missile guidance

    C. Navigation system updates

    D. Line-of-sight stabilization for:

        1. Angle rate bombing systems

        2. Air-to-air gunnery

        3. Reconnaissance sensors

        4. Target identification sensors

III. Ground Based Applications

    A. Range instrumentation systems for tracking aircraft and missiles with multiple cinetheodolite cameras

    B. Wind estimation in remote areas derived from cloud tracking in satellite imagery

    C. Anti-aircraft gun pointing systems

    D. Image registration for remote sensing using satellite imagery

2.1.1 Space Applications

In the space applications area, the theoretical basis for autonomous orbit determination using line-of-sight information from landmark-tracking systems has been established [28] . The recent Viking missions to Mars demonstrated the feasibility of using on-board imaging sensors

to aid in deriving the precision spacecraft tracking information required for accurate orbit insertion around the outer planets [20].

2.1.2 Airborne Applications

The airborne applications are the most highly developed real-time systems on the list. The Walleye electrooptically guided glide bomb, the GBU-8 and GBU-15 Modular Guided Glide Bombs (MGGB), and both the TV guided and imaging infrared versions of the AGM-65 Maverick missile use image-tracking systems to guide the weapons from launch to target [5],[22]. Other missiles use image-tracking systems in their terminal phase of flight to obtain extremely high accuracies [4]. Air-to-air image-tracking systems were tested as part of the AIM-82 program to develop a new air-to-air missile [6]. Problems were encountered when the trackers had difficulty maintaining stable tracking as the target flew in front of a cluttered background scene.

Examples of line-of-sight stabilization systems for target acquisition and identification are the Target Identification System Electrooptical (TISEO) employed on the F-4E and the Video Augmented Tracking System (VATS) modification to the PAVE TACK pod to provide automatic scene stabilization for the forward-looking infrared sensor in that system [22],[3],[23].

Another application which is being actively pursued is the use of range and line-of-sight rate information in lieu of a doppler radar to update inertial navigation system velocities. This requires very stable tracking of available landmarks. In a similar application, the U.S. Marine Corps A-4M will have a tracker on board to provide line-of-sight information for an angle-rate bombing system [23]. The use of

an image-tracking system is being investigated to determine the feasibility of using a director mechanization for air-to-air gunnery [33]. In this system, the tracker must stabilize the line-of-sight to the target.

2.1.3 Ground-Based Applications

Many ground-based applications of image-tracking systems are similar to the airborne and spaceborne applications in that they are real-time target tracking applications. Ground based processing of satellite imagery (such as earth resources data) is an exception. The difficulties of image registration from one satellite pass to another have been studied by Smith and Phillips [34]. This is an area where improvements in nonreal-time image-tracking appear to be usable. Another application for advanced tracking techniques is the estimation of wind speed and direction at remote locations. This has been demonstrated using satellite imagery [18].

2.1.4 Constraints On System Mechanization

The grouping of applications into space, airborne, and ground-based systems serves to divide the applications by both speed and equipment cost as well as by processing location. For the space applications, the hardware costs will be relatively high, and real-time processing will be required for those applications listed.

The airborne systems in the blowaway category (missile guidance units) must have a low unit cost and a high processing rate to accomplish their task. It is in this group that suboptimal performance may be tolerable to obtain the low cost and high speed required. The airborne scene stabilization systems can be more costly than the

17

throwaway guidance units but must operate at real-time rates, typically 30 or 60 images per second, and need a lower value of pointing-angle noise to accomplish their task of stabilizing long focal length telescopes. The ground-based applications for satellite image registration and range instrumentation can employ powerful and expensive large-scale computers and can operate at nonreal-time rates to obtain maximum accuracy. At the same time these systems are the slowest and the most expensive of the present image-tracking systems, yet with improved algorithms, further research into the underlying problems, and faster computers, these applications may eventually be carried out at real-time rates.

2.2 Classical Similarity Detection Techniques

Basic image-tracking techniques can be grouped into three categories:

1) Point-Tracking

2) Minimum-Distance-Tracking

3) Correlation-Tracking

2.2.1 Point Tracking

Point tracking proceeds on the assumption that objects to be tracked possess characteristics (features) which are not present in adjacent regions of the image. Under this assumption, a characteristic feature which identifies the target class is detected via a feature extraction process and used to derive the target position within the field of view. Sequential images are processed independently although adaptive feature extraction methods are usually employed. In their most

18

elementary form feature extractors have binary valued outputs, i.e. if a feature is present at a point in the image the output of the feature extractor is a one at that point, while, if the feature is not present the output is zero. Features which have been used include the following:

1) Intensity above a threshold

2) Intensity-derivative magnitude above a threshold

3) Intensity-derivative sign change

Examples of the results of these feature extractors operating on one-dimensional data are illustrated in Figure 3.

For single-feature point-trackers, error signals are computed using either an area-balance or centroid algorithm. The area-balance algorithm integrates the area of detected features in each quadrant of the gated region of the data image and forms an error signal for each axis by subtracting the resulting values in opposite halves of the gated region.

The so-called "centroid" algorithm is more often a first-moment algorithm than a true centroid computation. The first-moment of the area detected by the feature extractor is the usual error signal. Both the area balance and centroid error signal generating equations can be written in terms of a weighting function multiplied by the preprocessed data image value and integrated over the gated region. The horizontal and vertical error signals are

$$\epsilon_x = \iint_{\text{gated area}} \omega_x (x,y) \ f(x,y) \ dy \ dx \qquad (2.1)$$

and

19

INTENSITY

INTENSITY THRESHOLD

FIRST
DERIVATIVE
OF
INTENSITY

DERIVATIVE
MAGNITUDE
THRESHOLD
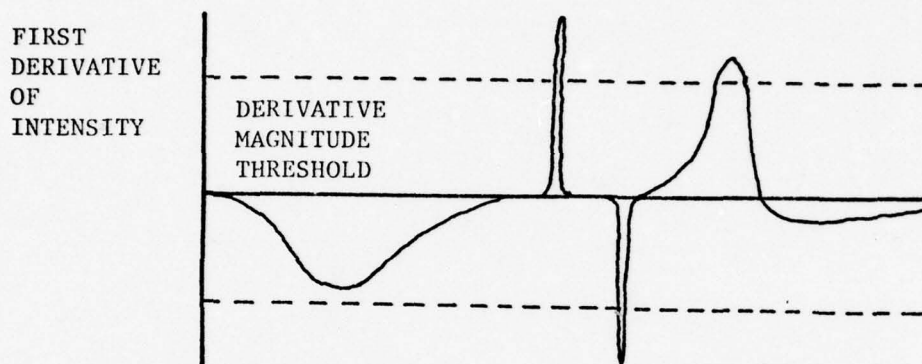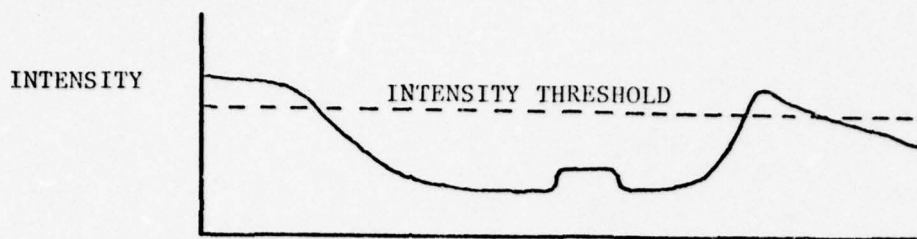
FEATURE EXTRACTOR OUTPUTS

INTENSITY
THRESHOLD

DERIVATIVE
MAGNITUDE
THRESHOLD

SIGN OF
DERIVATIVE

Figure 3.   Feature Extractors Operating on
One-Dimensional Data

20

PREPROCESSED DATA IMAGE
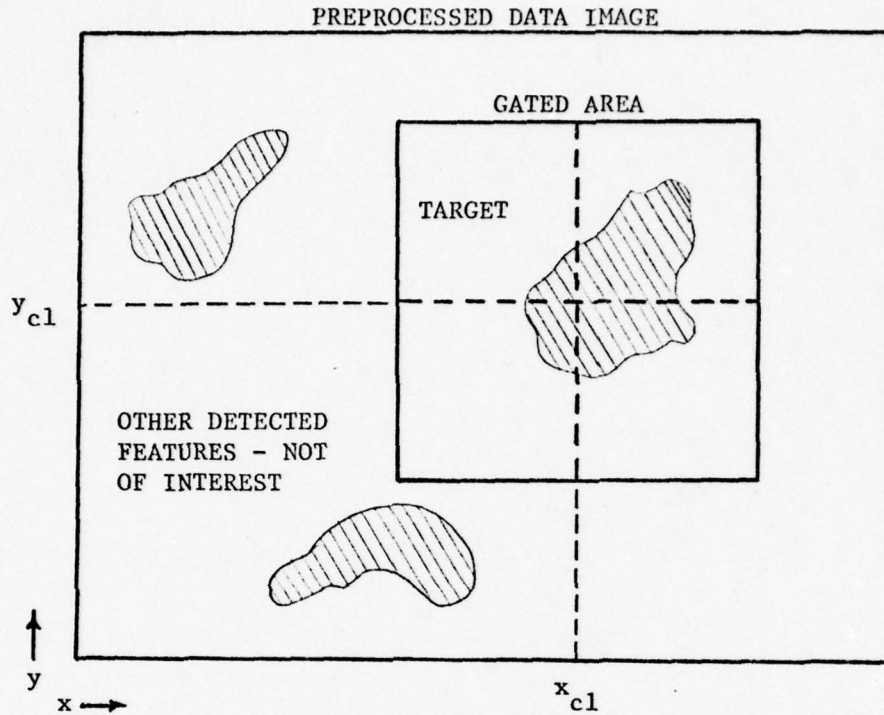
$$\mathbb{N} \quad f(x,y) = 1$$

$$\square \quad f(x,y) = 0$$

HORIZONTAL ERROR SIGNAL

$$\varepsilon_x = \iint \omega_x(x,y) f(x,y) dy dx$$

GATED AREA

VERTICAL ERROR SIGNAL

$$\varepsilon_y = \iint \omega_y(x,y) f(x,y) dy dx$$

GATED AREA

FOR AREA BALANCE

$$\omega_x(x,y) = SIGN(x-x_{c1}) \quad \omega_y(x,y) = SIGN(y-y_{c1})$$

FOR CENTROID

$$\omega_x(x,y) = x-x_{c1} \quad \omega_y(x,y) = y-y_{c1}$$

Figure 4. Error Computations for Point Tracking Algorithms

21

$$\epsilon_y = \iint\limits_{\text{gated area}} \omega_y(x,y) \, f(x,y) \, dy \, dx \tag{2.2}$$

where the weighting functions are:

for area balance

$$\omega_x(x,y) = \text{sign}(x - x_{cl}) \tag{2.3}$$

$$\omega_y(x,y) = \text{sign}(y - y_{cl}) \tag{2.4}$$

where $\quad \text{sign}(z) = \begin{cases} -1 & \text{for } z < 0 \\ 0 & \text{for } z = 0 \\ 1 & \text{for } z > 0 \end{cases}$

and for the centroid algorithm

$$\omega_x(x,y) = x - x_{cl} \tag{2.5}$$

$$\omega_y(x,y) = y - y_{cl} \tag{2.6}$$

and the notation is as shown in Figure 4.

A common method for using the error signal in point trackers is to drive the gate within the image to null the error signal and adjust the sensor pointing angles to center the gate within the field of view. Centroid trackers are quite susceptible to noise in the preprocessed data image when the detected target area is small compared to the gated area. Under this condition, a small region of noise at the edge of the gated area can cause an error of one quarter of the gate dimension

22

because of the relatively large magnitude of the moment contributed by the noise. This is one of the reasons why most of today's centroid trackers are mechanized with adaptive gates (Maverick, GBU-15, VATS). The gate size is decreased as the detected target area gets smaller in order to reduce the tracker's noise sensitivity. A variety of point-tracker exists which does not convert the intensity image to a binary-valued image but uses the intensity values themselves to compute an intensity centroid. This algorithm is limited to tracking targets which are significantly brighter (or darker) than their surroundings. The polarity of the incoming sensor data is reversed to track dark targets. A primary advantage of this arrangement is that there is no threshold to be set in the feature extractor.

Point-tracking algorithms are the basis for most of the current generation of real-time trackers for airborne applications. For these applications, the input signals are generally in the standard television format (525 or 875 line scan , either 2:1 or random interlace, 30 frames/second). The assumption that the selected feature is not present in the adjacent regions of the image guarantees only that the target feature is bounded and not that it is unique within the field of view. As a result, a subregion of the field of view must be gated for use by the tracker. A large number of different algorithms have been developed, some with adaptive gate size, some with fixed gate size, and some with multiple gates. Early algorithms used fixed feature extraction techniques; later, adaptive thresholds were employed. Error signal generation was by analog area-balance techniques for early designs, but first moment and true centroid algorithms have become standard in the most recent models. Selection of features (target

23

characteristics) for use in point-tracking algorithms has received substantial support from the Department of Defense. The two most common feature extraction processes are intensity thresholding and gradient thresholding, combined with elementary forms of spectral and spatial filtering. These forms of feature extraction have been selected in large part due to the ease with which they can be implemented with analog signal processing techniques.

2.2.2   Minimum Distance Tracking

In order to talk about the "distance" between two images as a measure of their similarity, it is necessary to represent images as elements of a vector space. While two-dimensional images are commonly visualized as two-dimensional arrays or matrices of intensity values, they can just as easily be thought of as vectors. Thus, the reference image, Ir, will have both a two-dimensional and a one-dimensional representation. Where the possibility of ambiguity exists, the bracketed notation will be used to indicate the matrix form.

$$
[Ir] = \begin{bmatrix} Ir(1,1) & \cdots & Ir(1,M_C) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ Ir(M_R,1) & \cdots & Ir(M_R,M_C) \end{bmatrix} \tag{2.7}
$$

24

$$Ir = \begin{bmatrix} Ir(1,1) \\ \cdot \\ \cdot \\ \cdot \\ Ir(M_R,1) \\ Ir(1,2) \\ \cdot \\ \cdot \\ \cdot \\ Ir(M_R,M_C) \end{bmatrix} \qquad (2.8)$$

where M_R and M_C are the number of elements in each row and column

respectively.

A function D is called a distance function if for a vector

space X, with x1, x2, and x3 in X, D satisfies the following conditions

[35]:

$$D(x1,x2) = D(x2,x1)$$
$$D(x1,x3) \leqslant D(x1,x2) + D(x2,x3)$$
$$D(x1,x2) = 0 \text{ if and only if } x1 = x2 \qquad (2.9)$$
$$D(x1,x2) \geqslant 0$$

It is possible to define legitimate distance functions in

unconventional ways and use these functions to measure the degree of

similarity between two images. One such distance function is a

relational metric which depends not on the intensity at each point in

the image but only on the relationship between intensity values at

adjacent points and in adjacent regions. By coding the components of

the vector as follows

25

RELATION	BINARY REPRESENTATION
<	0 1
=	0 0
>	1 0

the Hamming distance can be employed to measure similarity between images [36]. The importance of this idea is to recognize that unconventional distance functions may be useful if they improve performance or simplify computation.

The conventional distance functions for images are the norms of the difference image formed between two image vectors. The norm of a vector x will be indicated by $||x||$ and is related to a specific inner product by the formula

$$||x||^2 = (x , x) \qquad (2.10)$$

where $(.,.)$ denotes a generalized inner product [35]. The generalized inner product of two conformable vectors is defined by a positive definite Hermitian matrix [24]. If H is a positive definite Hermitian matrix, and x1 and x2 are column vectors, the inner product (x1,x2) defined by H is

$$(x1 , x2)_H = x1^T H x2 \qquad (2.11)$$

where T indicates the transpose of the vector. If H = I, the identity

26

matrix, then (.,.) is the familiar vector dot product. Minimum-norm tracking seeks to minimize the norm of a difference image formed by subtracting the data image from the reference image on a point-by-point basis. The two most common norms are

$$||x||_1 = \sum_{i=1}^{N} |x_i| \qquad (2.12)$$

and

$$||x||_2 = \left[\sum_{i=1}^{N} (x_i)^2 \right]^{\frac{1}{2}} \qquad (2.13)$$

These correspond to the Minkowski norms of order one and two, respectively [35]. Henceforth, when a norm is used without a subscript, it will be understood to be the Minkowski norm of order two with uniform weights ($H = I$).

Barnea and Silverman have developed a class of minimum distance algorithms for fast digital image registration called Sequential Similarity Detection Algorithms (SSDA's) [8]. These algorithms allow the use of any distance function which can be defined at each point in an image pair, and under conditions of high signal-to-noise ratios for the imagery being processed, require considerably less computation to find a minimum than the exhaustive search techniques used previously. Webber has described techniques for setting the SSDA threshold[38].

27

2.2.3   Correlation Tracking

Correlation trackers seek the location of the subset of the data image which maximizes either the cross correlation function or the normalized cross correlation function between a reference image and that particular sub-image. Let $Ir_L(i,j)$ be an L by L reference image, and let $Id(i,j)$ be an M_R by M_C data image. The elements of the unnormalized cross correlation surface $R_{rd}(u,v)$ are defined to be

$$R_{rd}(u,v) = \sum_{i=1}^{L}\sum_{j=1}^{L} Ir_L(i,j)\,Id(i+u,j+v) \qquad (2.14)$$

$$1 < u < M_R + 1 - L$$
$$1 < v < M_C + 1 - L$$

By finding the (u,v) which maximizes this function, the translational registration error is determined [8]. Normalization is accomplished by dividing this function by the product of the autocorrelation functions of the reference image and the data subimage. The normalized cross correlation surface is defined by

$$R_{rd}(u,v) = \frac{\left[\sum_{i=1}^{L}\sum_{j=1}^{L} Ir(i,j) Id(i+u,j+v)\right]^2}{\left[\sum_{i=1}^{L}\sum_{j=1}^{L} Ir(i,j)^2\right]\left[\sum_{i=1}^{L}\sum_{j=1}^{L} Id(i+u,j+v)^2\right]} \qquad (2.15)$$

These expressions can be rewritten as inner products in the following form:

$$R_{rd}(u,v) = (Ir, Id_{u,v}) \qquad (2.16)$$

$$R_{rd}(u,v) = \frac{(Ir, Id_{u,v})^2}{(Ir, Ir)(Id_{u,v}, Id_{u,v})} \qquad (2.17)$$

where $Id_{u,v}(i,j) = Id(u+i,v+j)$, and Ir is a suitable reference image.

Various preprocessing techniques have been investigated to enhance the raw imagery prior to performing the cross correlation. Hayes has suggested that thresholding, Laplacian enhancement, edge enhancement, and neighborhood averaging may be appropriate techniques to apply depending on the sensor used to obtain the imagery and the

imagery itself [13]. Cross correlation between transformed images has also been demonstrated with inversion [13], Fourier transformation (magnitude and phase correlation) [1], marginal summation [41], and phase correlation on the marginal sums [11]. Pratt has shown that the peak of the statistical correlation measure can be appreciably sharpened by application of linear spatial preprocessing [29].

2.3 Summary

In this chapter an historical summary of image tracking applications and similarity detection techniques was presented. In the following chapters an appropriate signal-to-noise ratio will be developed for the image tracking task, and four new techniques will be introduced to increase the effective tracking signal-to-noise ratio.

Chapter 3

SIMILARITY DETECTION

In this chapter, the relationship between minimum norm and maximum cross correlation techniques for similarity detection is shown, the characteristics of the minimum norm distance function as applied to a difference image are investigated, and a signal-to-noise ratio which is applicable to the minimum distance similarity detection problem is developed. In Chapter 4, signal-to-noise enhancement techniques are developed, and in Chapter 6 these techniques are applied to an integrated tracking algorithm.

3.1   Minimum Norm vs Maximum Cross Correlation

Similarity detection via minimization of the norm of the difference between two images is equivalent to maximizing the cross correlation function between the same images under a restricting set of assumptions. To see this, express the square of the norm of the difference between two images as an inner product.

$$||Ir-Id||^2 = (Ir-Id, Ir-Id) \qquad (3.1)$$

Expanding this expression yields

$$||Ir - Id||^2 = (Ir,Ir) - (Ir,Id) - (Id,Ir) + (Id,Id) \qquad (3.2)$$

and since the image components are real valued, the conjugate symmetry of the inner product allows us to write this as

$$||Ir - Id||^2 = (Ir,Ir) + (Id,Id) - 2(Ir,Id) \qquad (3.3)$$

Since Ir is the reference image, it is a constant for all trial data images. The last term on the right-hand side of (3.3) is the cross correlation between Ir and Id (see Section 2.2.3). From this we observe that if (Id,Id) is a constant, then minimizing the norm of the difference image is equivalent to maximizing the cross correlation function between Ir and Id. When (Id,Id) is a constant, the procedures which will be developed in the following sections with respect to minimum norm algorithms will produce results which are equivalent to maximum cross correlation tracking algorithms. In all other cases, while results may be similar, no guarantee is made about their equivalence.

All of the tracking algorithms to be investigated will be minimum norm algorithms. Specifically, with the exception of the non-uniformly weighted norm which will be developed in Section 4.1 all norms will be uniformly weighted Minkowski norms of order two.

3.2 Notation

In order to facilitate precise descriptions and a compact notation, the following list of terms and symbols will be useful:

UNDERLYING IMAGE, Is - A perfect representation of the scene being viewed by the sensor. Both Ir and Id are noise corrupted versions of Is.

32

$$I_s(i,j)$$

$$1 \leqslant i \leqslant M_R$$

$$1 \leqslant j \leqslant M_C$$

REFERENCE IMAGE, I_r – The stored reference image is the "best available" representation of the scene being viewed by the sensor. In this context, "best available" depends on the particular tracker, sensor, etc.. If the tracker uses a prestored reference, the reference image may be an old reconnaissance photograph, or a manually generated drawing. I_r is an $M_R \times M_C$ digitized image.

$$I_r(i,j)$$

$$1 \leqslant i \leqslant M_R$$

$$1 \leqslant j \leqslant M_C$$

DATA IMAGE, I_d – An $M_R \times M_C$ digitized image, taken directly from the sensor.

$$I_d(i,j)$$

$$1 \leqslant i \leqslant M_R$$

$$1 \leqslant j \leqslant M_C$$

33

REFERENCE INDEX SET, L - An Nx2 index set specifying the row and column associated with a particular element of the reference set. If the reference set was always going to be a contiguous block of pixels of a fixed size, then the location of one corner and the length of each side would suffice to identify all of the pixels in the reference set. Historically this has been the configuration for the reference set [40],[8]. In Section 4.3 we will see that the effective signal-to-noise ratio is increased by selecting pixels from high signal regions of the image for inclusion in the reference set. Because the pixels in the reference set may be spread out over the whole image, the reference index set is required to keep track of the row and column associated with each included pixel.

$$L(i,j)$$
$$1 \leqslant i \leqslant N$$
$$j = 1,2$$

REFERENCE SET - The subset of the reference image indexed by the reference index set.

$$Ir(L(i,1),L(i,2))$$
$$1 \leqslant i \leqslant N$$

DIFFERENCE IMAGE, D - The image formed by shifting $Id(.,.)$ with respect to $Ir(.,.)$ and taking the point-by-point difference. The difference image is not defined where Ir and the shifted Id do not overlap.

34

$D(dy,dx,i,j)$

$$-dy_{MAX} \leqslant dy \leqslant dy_{MAX}$$

$$-dx_{MAX} \leqslant dx \leqslant dx_{MAX}$$

$$\max[1,dy] \leqslant i \leqslant \min[M_R, M_R - dy]$$

$$\max[1,dx] \leqslant j \leqslant \min[M_C, M_C - dy]$$

COMPARISON SET, Ic – The set of difference-image elements specified by the reference index set.

$$Ic(dy,dx,i) \triangleq D(dy,dx,L(i,1),L(i,2)) \tag{3.4}$$

$$1 \leqslant i \leqslant N$$

$$-dy_{MAX} \leqslant dy \leqslant dy_{MAX}$$

$$-dx_{MAX} \leqslant dx \leqslant dx_{MAX}$$

TRIAL REGISTRATION – The relative shift of the data image with respect to the positions of the picture elements as received from the sensor. Each trial registration produces a different comparison set. The trial registration which produces the comparison set with minimum norm will be denoted $(\widehat{dy}, \widehat{dx})$, otherwise the trial registration will be denoted (dy, dx).

35

3.3 Assumptions

It is assumed that:

1) Id differs from Ir by a translational misregistration and additive zero-mean noise in each image

$$Id(i,j) - n_d(i,j) = Ir(i+\widehat{dy}, j+\widehat{dx}) - n_r(i+\widehat{dy}, j+\widehat{dx})$$

$$(3.5)$$

2) The noise components of Ir and Id are uncorrelated

$$E[n_d(i,j)n_r(i+dy, j+dx)] = 0 \qquad (3.6)$$

where $n_r(.,.)$ is the zero mean noise associated with the reference image

$n_d(.,.)$ is the zero mean noise associated with the data image, and

$E(.)$ is the expected value operator.

Both of these assumptions will hold when the principle source of noise is electronic shot noise, and the average scene illumination changes slowly with respect to the frame rate.

3.4 Distance Function

The distance function $d_A^2(dy, dx)$ is defined as the weighted norm of $Ic(dy, dx, .)$

36

$$d^2_A(dy,dx) = Ic(dy,dx,.)^T \; A \; Ic(dy,dx,.)$$

$$= \sum_{i=1}^{N} a_{ii} \; Ic(dy,dx,i)^2 \qquad (3.7)$$

where A is a diagonal matrix of positive weighting factors with elements a_{ij}. When A is the identity matrix, we will use the notation $d^2_I(dy,dx)$ to indicate the specific case of equal weights for all components of $Ic(dy,dx,.)$.

3.5 Tracking Algorithm

The tracking algorithm will compute $d^2_A(dy,dx)$ for a range of dy and dx values, and select the (dy,dx) which corresponds to the minimum value of $d^2_A(dy,dx)$ as the relative translation of Id with respect to Ir. The range of values for dy and dx specifies the size of the search region. All search regions will be treated as being symmetric with $-dy_{MAX} \leq dy \leq dy_{MAX}$ and $-dx_{MAX} \leq dx \leq dx_{MAX}$. This is based on the assumptions that the search is centered on the most likely location for the image, and that the distribution of errors is symmetric about that location.

37

3.6 Noise Characteristics

The reference image Ir is of course not a perfect representation of the underlying scene being viewed by the sensor. If Ir is obtained directly from the sequence of data images, it contains noise with the same characteristics as all of the other data images. If the reference image is obtained by filtering the sequence of data images, then the noise component of Ir will differ from that of the raw data images. The variance of the reference image noise component associated with the image location Ir(i,j) will be denoted by $\sigma^2_{n,REF}(i,j)$, that is

$$\sigma^2_{n,REF}(i,j) = E[\, n_r^2(i,j)\,] \qquad (3.8)$$

The noise variance associated with the (i,j) coordinate of the data image Id is $\sigma^2_{n,DATA}(i,j)$.

$$\sigma^2_{n,DATA}(i,j) = E[n_d^2(i,j)] \qquad (3.9)$$

For sensors with single channel outputs where thermal noise is the dominant component of n_d, we will assume that the noise is ergodic and hence has stationary statistics [27]. Vidicons, laser line scanning systems with single detectors, and certain infrared scanners fall into this class of imaging sensor [39]. In addition, since the

38

sensor output is generally low pass filtered prior to digitization, a first order Markov model will be used for the sensor noise statistics. In Section 3.8 a model for imagery will be developed and the Markov nature of the sensor noise will be seen.

Stationary noise statistics allow us to eliminate the spatial specificity of the noise variances and simply write them as $\sigma^2_{n,REF}$ and $\sigma^2_{n,DATA}$. Spatial and temporal ergodicity will allow us to estimate the noise variance at an arbitrary point in the image from the sample variance of the noise over the whole image (note that for scanned sensors spatial and temporal ergodicity are equivalent).

3.7 Characteristics of the Auto-Distance Function

The auto-distance function is the two-dimensional distance function which results from computing $d_A^2(dy,dx)$ between an image and a translated copy of the same image. The auto-distance function has characteristics which are determined by the statistical signal and noise properties of the image itself.

Consider the case of shifting the reference image with respect to itself in a direction parallel to a scan line from an initially registered position. At registration, the distance $d_I^2(0,0)$ is zero since all of the elements of the comparison set are zero.

$$Ic(0,0,i) = D[0,0,L(i,1),L(i,2)]$$

$$= Ir[L(i,1),L(i,2)] - Ir[L(i,1)+0,L(i,2)+0]$$

$$= 0 \tag{3.10}$$

For a shift of one picture element, or pixel, $d_I^2(0,1)$ has contributions from both signal and noise components.

$$Ic(0,1,i) = Is[L(i,1),L(i,2)]+n_r[L(i,1),L(i,2)]$$

$$-Is[L(i,1),L(i,2)+1]-n_r[L(i,1),L(i,2)+1] \tag{3.11}$$

40

Let

$$m_i(dy,dx) = Is[L(i,1)+dy, L(i,2)+dx] - Is[L(i,1), L(i,2)] \qquad (3.12)$$

then

$$d_A^2(dy,dx) = \sum_{i=1}^{N} a_{ii}\left[m_i(dy,dx) + n_r[L(i,1), L(i,2)] \right.$$

$$\left. - n_r[L(i,1)+dy, L(i,2)+dx] \right]^2 \qquad (3.13)$$

The signal component of $d_A^2(dy,dx)$ is the contribution from $\{m_i(dy,dx)\}$. The noise terms represent the difference between the noise at two pixel locations separated from each other by a translation of (dy,dx). For line scanned imagery, the noise will be modeled as a Markov process in time; thus the noise will be correlated with itself much more strongly in the direction of scan than in the direction perpendicular to the scan.

Taking the expected value of (3.13) yields

41

$$E[d^2_A(dy,dx)] = \sum_{i=1}^{N} a_{ii} m^2_i(dy,dx)$$

$$+ \sum_{i=1}^{N} a_{ii} E\left[\left(n_r[L(i,1),L(i,2)]\right.\right.$$

$$\left.\left. - n_r[L(i,1)+dy,L(i,2)+dx]\right)^2\right] \tag{3.14}$$

and by writing the noise contribution in terms of the reference image

noise variance $\sigma^2_{n,REF}$ and the correlation coefficient for the noise

$\rho(dy,dx)$, we have an expression for the auto-distance function of the

reference image

$$E[d^2_A(dy,dx)] = \sum_{i=1}^{N} a_{ii} m^2_i(dy,dx)$$

$$+ 2 \sigma^2_{n,REF} [1- \rho(dy,dx)] \sum_{i=1}^{N} a_{ii} \tag{3.15}$$

Using uniform weights and taking an average over the N elements in Ic

yields

42

$$E\left[\frac{d_I^2(dy,dx)}{N}\right] = m_{N,AVE}^2(dy,dx) + 2\,\sigma_{n,REF}^2\,[1-\rho(dy,dx)] \qquad (3.16)$$

where

$$m_{N,AVE}^2(dy,dx) = \frac{1}{N}\sum_{i=1}^{N} m_i^2(dy,dx) \qquad (3.17)$$

Let $\mathcal{T}(dy,dx)$ be the scan time difference between two pixels separated by (dy,dx)

$$\mathcal{T}(dy,dx) \triangleq \mathcal{T}_y\,dy + \mathcal{T}_x\,dx \qquad (3.18)$$

where $\mathcal{T}_y$ is the time that the sensor takes to scan one entire line and $\mathcal{T}_x$ is the scan time between two adjacent pixels in the same line.

For a first order Markov process (see Section 3.6) the normalized autocorrelation function has the form [12]

$$\phi(\mathcal{T}) = e^{-\alpha|\mathcal{T}|} \qquad (3.19)$$

where $1/\alpha$ is the correlation time of the process. The correlation coefficient $\rho(dy,dx)$ in (3.16) is the value of the normalized autocorrelation function of the noise when the time delay is the scan time difference between two pixels separated by (dy,dx). Substituting (3.18) into (3.19) yields an expression for the correlation coefficient in terms of dy and dx.

43

$$\rho(dy, dx) = \exp\left[-\alpha\left|\mathcal{T}_y dy + \mathcal{T}_x dx\right|\right] \qquad (3.20)$$

For a minimum resolution TV compatible scanner with 256 pixels/horizontal line ($\mathcal{T}_y > 256\mathcal{T}_x$) and a correlation time of a few pixels (say $1/\alpha < 4\mathcal{T}_x$), it is clear that $\rho(dy,dx)$ is approximately zero for $dy \neq 0$.

$$\rho(dy, dx) < \exp\left[-\frac{1}{4\mathcal{T}_x}\left|256\,\mathcal{T}_x\,dy + \mathcal{T}_x\,dx\right|\right]$$

$$< \exp\left[-64\left|dy\right| - .25\left|dx\right|\right] \qquad (3.21)$$

Thus we can assume that the noise component of line scanned imagery is uncorrelated between pixels which are adjacent to each other in a direction perpendicular to the direction of scan, but we must take into account the correlation which exists between pixels that are adjacent to each other in the direction of scan.

3.8   A Facet Model for Imagery

In order to do anything useful with $d^2_A(dy,dx)$, we need an image model with tractable characteristics for $m(dy,dx)$. The model that we will use assumes that the image intensity is a well-behaved function that can be approximated by a local tangent plane (facet) in the

44

vicinity of each pixel coordinate. The resulting expression for $Is(i+dy, j+dx)$ is a truncated Taylor series in two variables

$$Is(i+dy, j+dx) = Is(i,j) + dy \, \nabla Is(i,j) \cdot \hat{1}_y$$

$$+ \, dx \, \nabla Is(i,j) \cdot \hat{1}_x \qquad (3.22)$$

where ∇ is a discrete gradient operator defined by

$$\nabla Is(i,j) = \frac{Is(i+1,j) - Is(i-1,j)}{2} \hat{1}_y$$

$$+ \frac{Is(,j+1) - Is(i,j-1)}{2} \hat{1}_x \qquad (3.23)$$

and $\hat{1}_x$ and $\hat{1}_y$ are unit vectors parallel to the x and y axes respectively.

For $L(k,1)=i$ and $L(k,2)=j$ the expression for $m_k(dy,dx)$ is

$$m_k(dy,dx) = \nabla Is(i,j) \cdot (dy \, \hat{1}_y + dx \, \hat{1}_x) \qquad (3.24)$$

and substituting (3.24) into (3.17) the general expression for $m^2_{N,AVE}(dy,dx)$ is

45

$$m^2_{N,AVE}(dy,dx) = \frac{1}{N}\sum_{i=1}^{N}\left[\nabla Is[L(i,1),L(i,2)]\cdot(dy\,\hat{1}_y + dx\,\hat{1}_x)\right]^2$$

$$(3.25)$$

Figure 5 shows the expected value of the normalized auto-distance function for a signal-free image ($m^2_{N,AVE}(dy,dx)=0$) with a noise variance of $\sigma^2_n=3$ and a noise-correlation coefficient of the form

$$\rho(dy,dx) = \begin{cases} \exp[-.7|dx|] & \text{for } dy = 0 \\ \\ 0 & \text{otherwise} \end{cases} \qquad (3.26)$$

Figure 6 shows the value of the auto-distance function for a noise-free image ($\sigma^2_n=0$) with average signal strength

$$m^2_{N,AVE}(dy,dx) = 1.66\,dy^2 + .16\,dx^2 \qquad (3.27)$$

where the coefficients are adjusted to match the empirical data listed in Table 1. Figure 7 shows the expected value of the auto-distance function for an image which combines the noise characteristics from Figure 5 and the signal characteristics from Figure 6.

To demonstrate the facet image model by comparison with

46

$$2\sigma^2_n = 6$$
$$\rho = 1-e^{-.7|dx|}$$
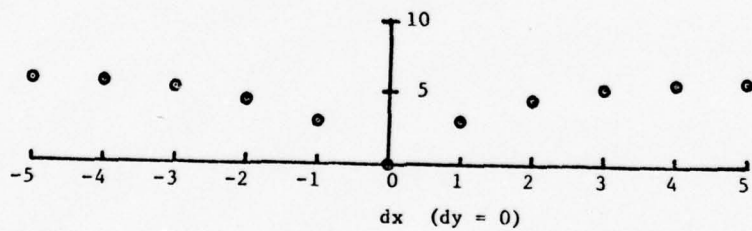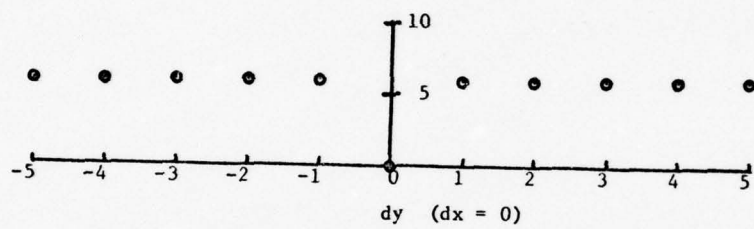
Figure 5. Auto-Distance Function for Pure Noise

47
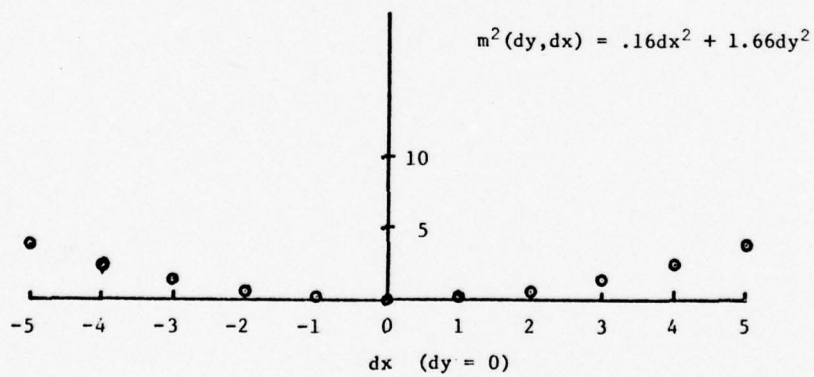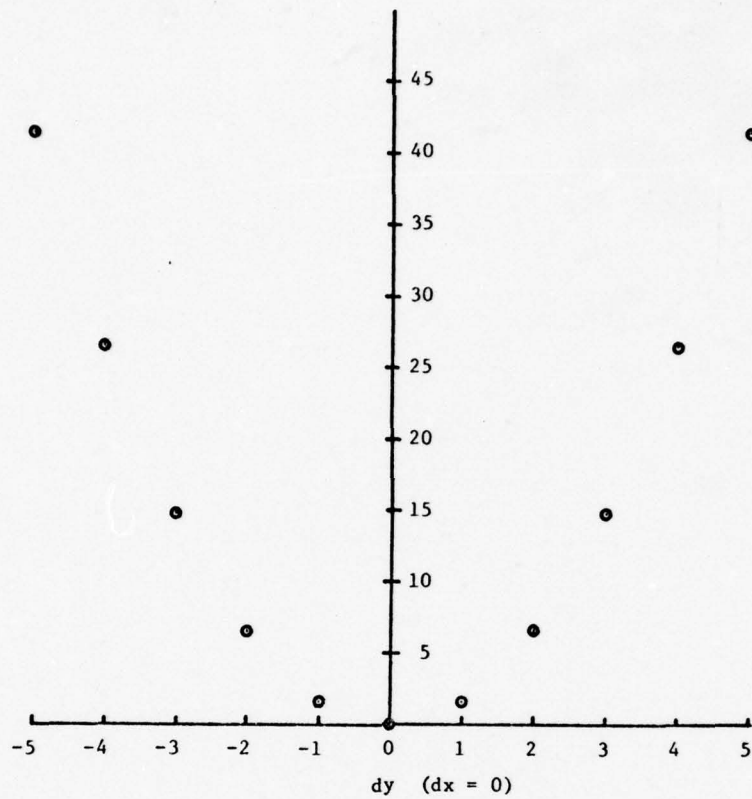
Figure 6. Auto-Distance Function for a Simulated
Noise-Free Image

48

$$2\sigma_n^2 = 6 \qquad \rho = 1-e^{-.7|dx|}$$

$$m^2(dy,dx) = .16dx^2 + 1.66dy^2$$
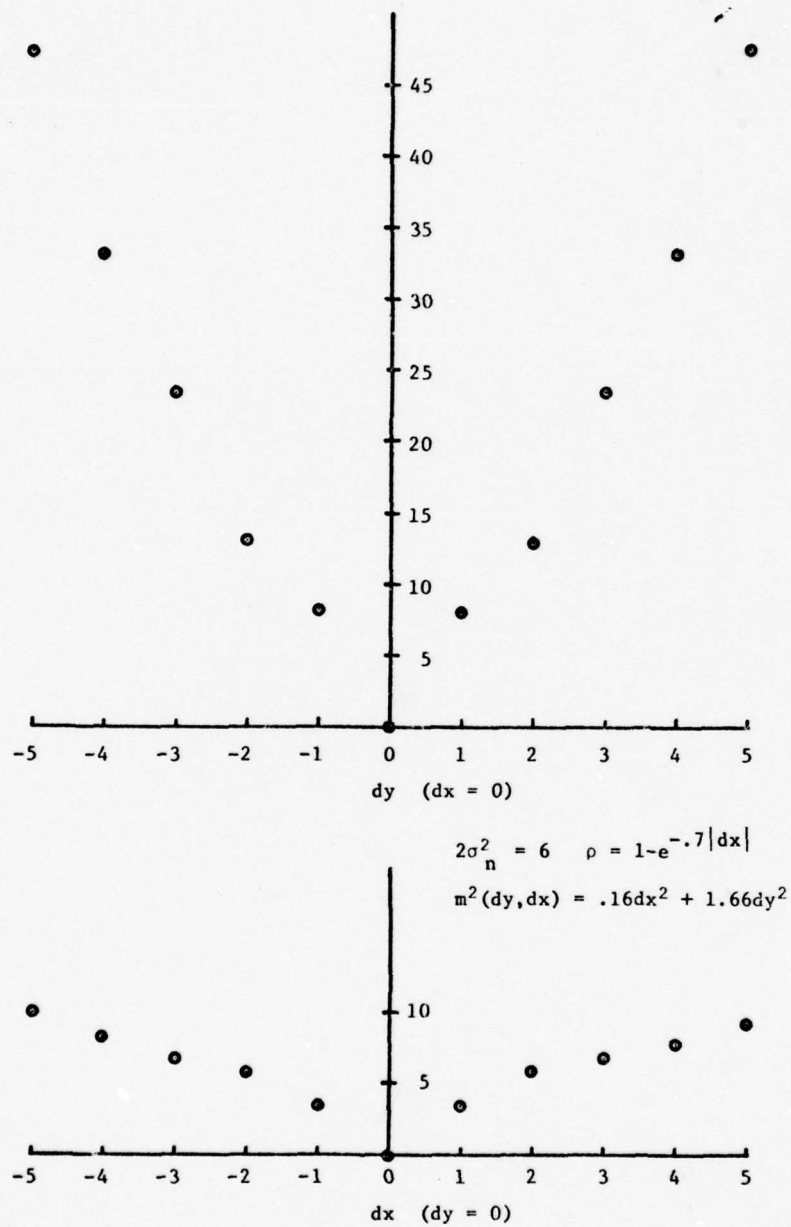
Figure 7. Auto-Distance Function for a Simulated Noisy Image

49

experimental data, Table I lists the values of the auto-distance

function for a small subimage taken from a frame of television imagery

(see Appendix I for source data). Figure 8 shows the plotted

values for dy = 0 and dx = 0.

Note the markedly smaller signal component in the direction

parallel to the axis of the scan (x-axis) and the similarity between

Figure 7 and Figure 8. The differential signal strength in the x and y

directions is attributable to a high digitizer sampling rate relative

to the bandwidth of the output signal from the sensor. The similarity

of Figure 7 and Figure 8 demonstrates the validity of the facet model

and the Markov nature of the noise.

3.9   Cross-Distance Function

When $d_I^2(dy, dx)$ is used to compare a reference image Ir with

noise variance $\sigma_{n,REF}^2$ to a data image Id with noise variance $\sigma_{n,DATA}^2$,

the resulting two-dimensional distance function will be called the

cross-distance function. The cross-distance function is biased with

respect to the auto-distance function by an amount equal to the sum of

the noise variances from the reference image and the data image. For a

correctly registered data image

Table I. Auto-distance function values for frame 1 of AIRPLANE

						dx					
dy	-5	-4	-3	-2	-1	0	1	2	3	4	5
-5	48.35	47.57	47.24	46.76	46.64	47.29	47.79	47.89	48.51	49.60	50.20
-4	35.81	34.41	33.54	33.12	32.82	33.15	33.89	34.13	34.69	35.86	36.83
-3	28.18	26.10	24.88	23.89	23.16	23.54	24.42	24.87	25.62	27.10	28.49
-2	18.50	16.15	14.75	13.66	13.12	13.03	13.61	14.21	14.97	16.54	18.59
-1	14.52	12.20	10.62	9.53	8.63	8.17	8.95	10.14	11.02	12.45	14.44
0	9.99	8.16	6.84	5.82	3.44	0.00	3.47	5.94	6.86	7.92	9.16
1	13.76	11.68	10.45	9.75	8.76	8.06	8.59	9.35	9.94	11.02	12.58
2	17.39	15.45	14.41	13.96	13.52	12.98	13.16	13.23	13.41	14.04	15.19
3	26.78	25.70	25.01	24.77	24.34	23.54	23.32	23.33	23.16	23.35	24.06
4	35.58	34.71	34.27	34.30	34.10	33.19	32.90	32.37	31.60	31.17	31.19
5	48.98	48.77	48.64	48.61	48.30	47.50	47.05	46.29	45.20	44.09	43.58
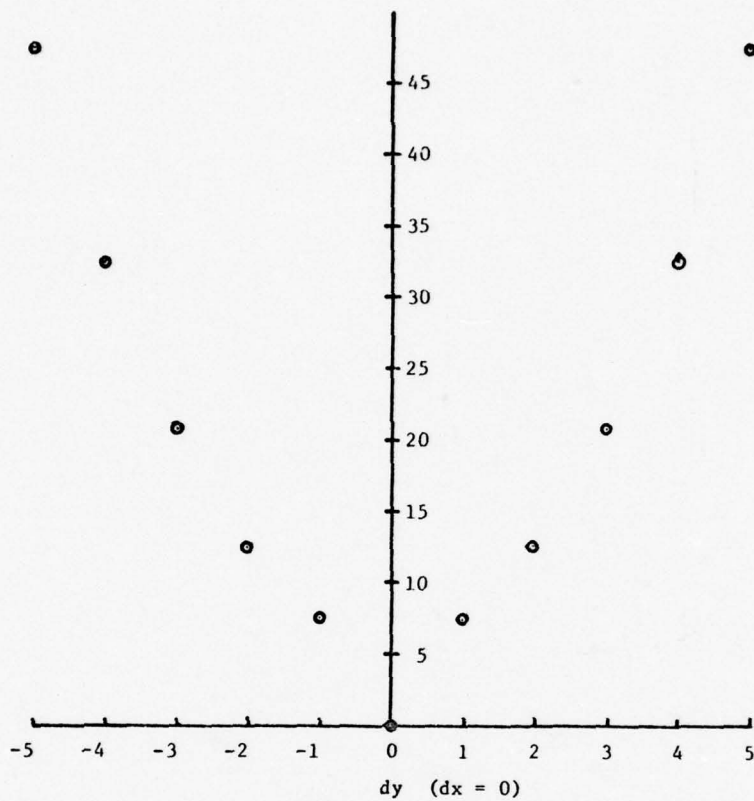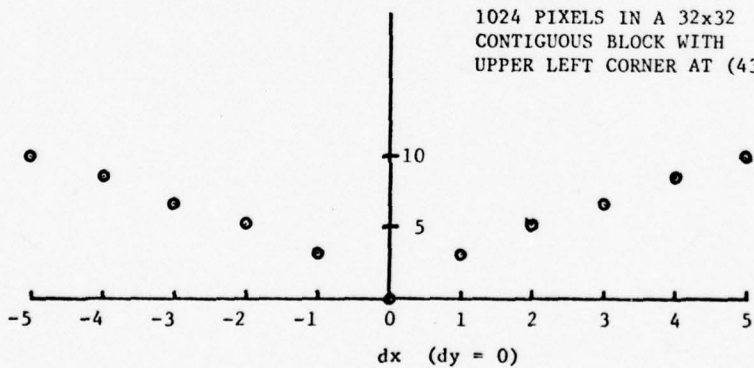
Figure 8. Auto-Distance Function for Frame 1
of Image Sequence AIRPLANE

52

$$E\left[\frac{d_I^2(0,0)}{N}\right] = \frac{1}{N}E\left[\sum_{i=1}^{N} Ic(0,0,i)^2\right]$$

$$= \sigma_{n,REF}^2 + \sigma_{n,DATA}^2 \qquad (3.28)$$

and for a relative translation of (dy,dx)

$$E\left[\frac{d_I^2(dy,dx)}{N}\right] = m_{N,AVE}^2(dy,dx) + \sigma_{n,REF}^2 + \sigma_{n,DATA}^2 \qquad (3.29)$$

Table II lists the values of the cross-distance function between two successive frames of actual T.V. imagery, and Figure 9 shows the plotted values for dy = 0 and dx = 1. In this case, the minimum distance match occurs at a translation of (0,1). It is the location of this minimum distance coordinate that the tracking algorithm declares to be the present position of the current reference image. If the location of the minimum distance coordinates change on a frame-to-frame basis, the change is interpreted to be either sensor motion or image motion.

In the next two sections we will develop an expression for the probability of making a particular error in determining the correct registration of a data image with respect to the reference image, propose a natural "signal-to-noise" ratio for the minimum norm tracking problem and show the relationship between this signal-to-noise ratio and the probability of error.

53

Table II. Cross-distance function values for frames 1 and 2 of image sequence AIRPLANE

| | | | | | | dx | | | | | |
dy	-5	-4	-3	-2	-1	0	1	2	3	4	5
-5	44.20	43.68	43.30	43.05	43.06	43.11	43.56	44.30	44.96	45.39	46.21
-4	36.54	35.53	34.67	33.72	33.46	33.39	33.59	34.35	35.12	35.59	36.77
-3	26.72	25.05	23.53	22.37	21.73	21.74	22.14	22.75	23.27	23.82	25.01
-2	21.46	19.90	18.29	16.62	15.66	15.21	15.25	16.04	16.70	17.39	18.67
-1	15.28	13.24	11.38	9.84	9.02	8.34	7.89	8.45	9.16	10.21	11.66
0	14.94	12.92	10.93	9.24	8.07	6.95	6.71	7.16	8.21	8.94	10.12
1	14.80	13.06	11.38	9.86	8.52	7.83	8.04	8.26	8.52	8.91	9.80
2	21.08	19.72	18.77	17.75	17.00	16.04	15.48	15.62	16.26	16.49	17.12
3	26.78	25.82	25.30	24.94	24.43	23.67	23.37	23.43	23.49	23.33	23.37
4	38.46	37.71	37.13	36.58	36.39	36.54	36.78	36.10	35.46	34.62	34.39
5	47.87	47.46	47.35	47.56	47.69	47.95	47.88	46.88	46.09	45.08	44.24
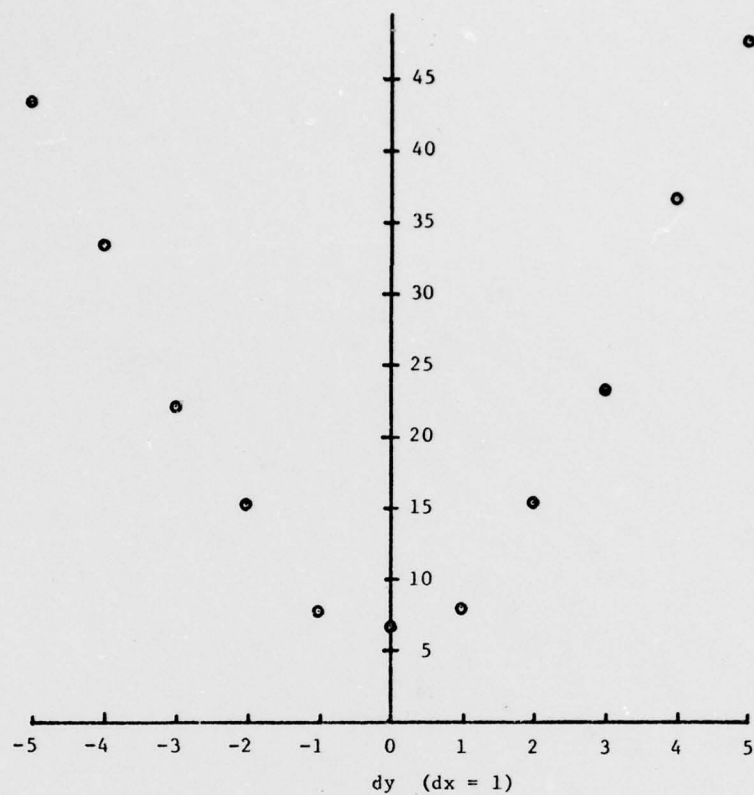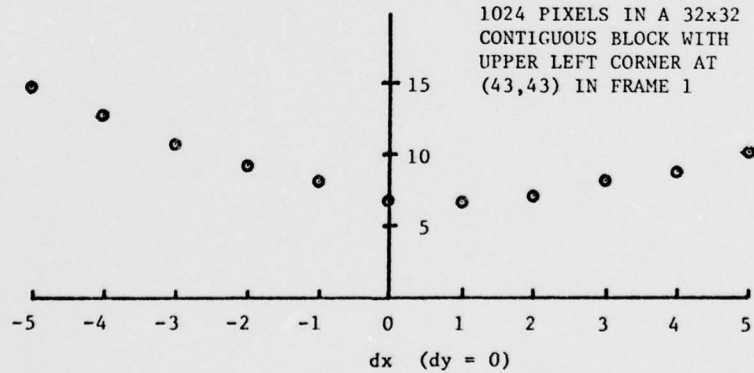
Figure 9. Cross-Distance Function for Frames 1 and 2 of Image Sequence AIRPLANE

55

3.10   Distance Function Statistics

Since each component of Ic is the difference between two noise-corrupted image intensity values, Ic is a random vector. The mean value of the ith component is

$$E[Ic(dy,dx,i)] = -m_i(dy,dx) \qquad (3.30)$$

and the variance is

$$var[Ic(dy,dx,i)] = var(n_r) + var[n_d]$$

$$= \sigma_{n,REF}^2 + \sigma_{n,DATA}^2 \qquad (3.31)$$

If n_r and n_d are Gaussian random variables (we will see later that this assumption is in good agreement with experimental data) then

$$\frac{d_I^2(dy,dx)}{\sigma_{n,REF}^2 + \sigma_{n,DATA}^2}$$

is a noncentral chi-square distributed random variable with noncentrality parameter Θ and N degrees of freedom where [16]

56

$$\Theta(dy,dx) = \sum_{i=1}^{N} \frac{m_i^2(dy,dx)}{\sigma_{n,REF}^2 + \sigma_{n,DATA}^2}$$

$$= \frac{Nm_{N,AVE}^2(dy,dx)}{\sigma_{n,REF}^2 + \sigma_{n,DATA}^2} \qquad (3.32)$$

For (dy,dx)=(0,0) the resulting distribution is central chi-square with N degrees of freedom.

For each (dy,dx), there exists the possibility that as a result of noise in both the reference image and the data image, the distance function value at the correct registration is greater than the distance function value at an incorrect registration. For a correctly registered data image the tracking algorithm will make an error any time that there is some (dy',dx') such that

$$d_I^2(0,0)-d_I^2(dy',dx') < 0 \qquad (3.33)$$

Under what conditions will this occur, and how is it related to the noise parameters? To explore these questions we let

$$\sigma_n^2 = \sigma_{n,REF}^2 + \sigma_{n,DATA}^2 \qquad (3.34)$$

and define two normalized random variables U and V

$$U(dy,dx) = \frac{d_I^2(dy,dx)}{\sigma_n^2} \qquad (3.35)$$

and

$$V = \frac{d_I^2(0,0)}{\sigma_n^2} \qquad (3.36)$$

In general U and V are not independent random variables since it is possible to have a reference index set with

$$[L(i,1),L(i,2)] = [L(j,1)+dy,L(j,2)+dx] \qquad (3.37)$$

for some legitimate set of (i,j,dy,dx). In fact, this case predominates when the reference set is composed of a contiguous block of pixels. Under this condition, a noise sample may contribute to both $d_I^2(0,0)$ and $d_I^2(dy,dx)$. The only way to assure that $d_I^2(0,0)$ and

58

$d_I^2 (dy,dx)$ are independent is to require that for all $i,j < N$

$$\left| L(i,1) - L(j,1) \right| > dy_{MAX} \qquad (3.38)$$

and

$$\left| L(i,2) - L(j,2) \right| > dy_{MAX} \qquad (3.39)$$

Figure 10 illustrates the spacing required to assure that U and V are uncorrelated when the noise samples are independent in a uniformly spaced reference set with $dy_{MAX} = dx_{MAX} = 5$ and $N = 64$.

Let $P_c (dy,dx)$ denote the probability that for a correctly registered data image $d_A^2 (0,0)$ is less than $d_A^2 (dy,dx)$

$$P_c (dy,dx) = P[d_A^2 (0,0) - d_A^2 (dy,dx) < 0] \qquad (3.40)$$

$P_c (dy,dx)$ is the probability of being correct with respect to the decision on whether the reference image is registered at $(0,0)$ or at (dy,dx). The probability of error, $P_\varepsilon (dy,dx)$ is the complement of $P_c (dy,dx)$

$$P_\varepsilon (dy,dx) = 1 - P_c (dy,dx) \qquad (3.41)$$

59

SINGLE PIXEL
SEARCH REGION

$N = 64$

$dy_{max} = dx_{max} = 5$

$2dx_{max} + 1$

$2dy_{max} + 1$

$dy_{max} + 1$

$|dx_{max} + 1|$
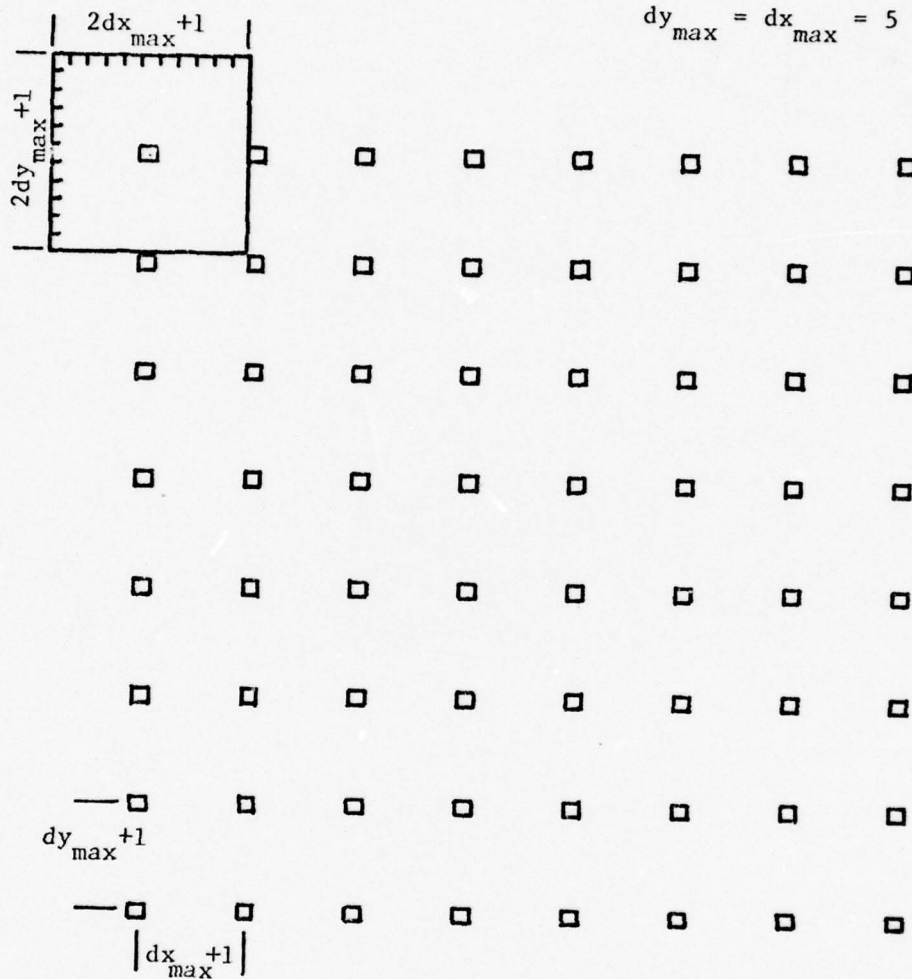
Figure 10. Relative Positions Specified by a
Reference Index Set for U and V to
Contain Uncorrelated Noise

60

Let

$$W(dy, dx) = V - U(dy, dx) \qquad (3.42)$$

For large values of N, both U and V tend toward a normal distribution [16]. We will normalize W(dy,dx) to zero mean and unit variance by the transformation

$$X(dy, dx) = \frac{W(dy, dx) - E\left[W(dy, dx)\right]}{\sqrt{var\left[W(dy, dx)\right]}} \qquad (3.43)$$

and for large N, approximate the probability distribution function of X by the unit normal distribution. Thus

$$\begin{aligned}
P_C(dy, dx) &= P\left[W(dy, dx) < 0\right] \\[2em]
&= erf^*\left\{\frac{-E\left[W(dy, dx)\right]}{\sqrt{var\left[W(dy, dx)\right]}}\right\} \\[2em]
&\triangleq erf^*\left[\mathcal{E}_N(dy, dx)\right] \qquad (3.44)
\end{aligned}$$

where $erf^*(\cdot)$ is defined by

$$erf^*(X) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X} exp\left[\frac{-t^2}{2}\right] dt \qquad (3.45)$$

61

We now derive the expressions for the argument of erf* (.) in terms of N and Θ , the parameters of the noncentral chi-square distributions associated with U and V.

$$E[W(dy,dx)] = E[V] - E[U(dy,dx)]$$

$$= N - [N + \Theta(dy,dx)]$$

$$= - \Theta(dy,dx) \qquad (3.46)$$

$$var[W(dy,dx)] = var[V - U(dy,dx)]$$

$$= E[V^2] + E[U^2(dy,dx)]$$

$$+ \Theta^2(dy,dx) + 2\Theta(dy,dx)E[V]$$

$$-2 E[VU(dy,dx)]$$

$$-2\Theta(dy,dx)E[U(dy,dx)] \qquad (3.47)$$

$$E[V] = N \qquad (3.48)$$

$$E[V^2] = N^2 + 2N \qquad (3.49)$$

62

$$E[U(dy, dx)] = N + \Theta(dy, dx) \qquad (3.50)$$

$$E[U^2(dy, dx)] = [N + \Theta(dy, dx)]^2 + 2[N + 2\Theta(dy, dx)] \qquad (3.51)$$

$$E[VU(dy, dx)] = \sigma_V \sigma_{U(dy, dx)} \rho_{VU} + E[V]E[U(dy, dx)]$$

$$= \rho_{VU} \sqrt{4N[N+2\Theta(dy, dx)]} + N[N+\Theta(dy, dx)]$$

$$(3.52)$$

where ρ_{VU} is the correlation coefficient between V and U(dy, dx).

Substituting (3.48) through (3.52) into (3.47) yields

$$\text{var}[W(dy, dx)] = 4N + 4\Theta(dy, dx)$$

$$- 4\rho_{VU}\sqrt{N^2 + 2N\Theta(dy, dx)}$$

$$(3.53)$$

63

and now substituting (3.32) for Θ, the argument of (3.44) becomes

$$\mathcal{E}_N(dy,dx) = \cfrac{\cfrac{N m_{N,AVE}^2(dy,dx)}{\sigma_n^2}}{\sqrt{4N + \cfrac{4N m_{N,AVE}^2(dy,dx)}{\sigma_n^2}} - 4\rho_{VU}\sqrt{2N + \cfrac{2N^2 m_{N,AVE}^2(dy,dx)}{\sigma_n^2}}}$$

$$= \cfrac{\sqrt{N}\,\cfrac{m_{N,AVE}^2(dy,dx)}{\sigma_n^2}}{2\sqrt{1 + \cfrac{m_{N,AVE}^2(dy,dx)}{\sigma_n^2}} - \rho_{VU}\sqrt{1 + \cfrac{2m_{N,AVE}^2(dy,dx)}{\sigma_n^2}}}$$

$$(3.54)$$

64

Let

$$\gamma^2_{N,AVE}(dy, dx) \triangleq \frac{m^2_{N,AVE}(dy, dx)}{\sigma^2_n} \qquad (3.55)$$

It appears that $\gamma^2_{N,AVE}(dy, dx)$ is the natural signal-to-noise ratio for

this minimum-norm detection problem. Notice that $\gamma^2_{N,AVE}(dy, dx)$ is

different for each (dy, dx). This signal-to-noise ratio is a two

dimensional function!

An upper bound can be established for $P_\mathcal{E}(dy, dx)$, the

probability of a particular error, by letting $\rho_{VU} = 0$ in (3.54). Then

$$\overline{\mathcal{E}}_N(dy, dx) = \frac{\sqrt{N}' \; \gamma^2_{N,AVE}}{2\sqrt{1 + \gamma^2_{N,AVE}}} \qquad (3.56)$$

and

$$P_\mathcal{E}(dy, dx) < 1 - \mathrm{erf} \; [\overline{\mathcal{E}}^*_N(dy, dx)] \qquad (3.57)$$

The upper bound is a function of $\gamma^2_{N,AVE}$ and N only. Figure 11 shows how the maximum probability of error varies with $\gamma^2_{N,AVE}$ for various values of N.

3.11 Experimental Probability of Error Determination

In order to assess the accuracy of the probability of error bound of (3.57) and to investigate the improvement that might be obtained when the noise component of the data image is correlated between trial registrations, a series of simulations was performed. For each simulation, the reference image was noise-free and noisy data images were generated by adding artificial Gaussian noise with variance σ^2_n to a copy of the reference image. Since the reference image is assumed "perfect", the auto-distance function provided a direct measure of m^2_{AVE} (dy,dx). Thus, for any reference index set and any trial registration, the signal-to-noise ratio was known. The probability of error was estimated by counting the fraction of the total number of trials on which d^2_I (dy,dx) was less than d^2_I (0,0). Figure 12 and Figure 13 present the results of two Monte Carlo simulation runs of 100 trials each. The 100 trials provide 95% confidence intervals ranging from $\pm.1$ for $P_\varepsilon = .5$ to $^{+.05}_{-.001}$ for P_ε near 0 [14].

$$P_\varepsilon < 1 - \text{erf} \star [\varepsilon_N]$$

$$\varepsilon_N = \frac{\sqrt{N}\, \gamma^2_{N,AVE}}{2\sqrt{1 + \gamma^2_{N,AVE}}}$$
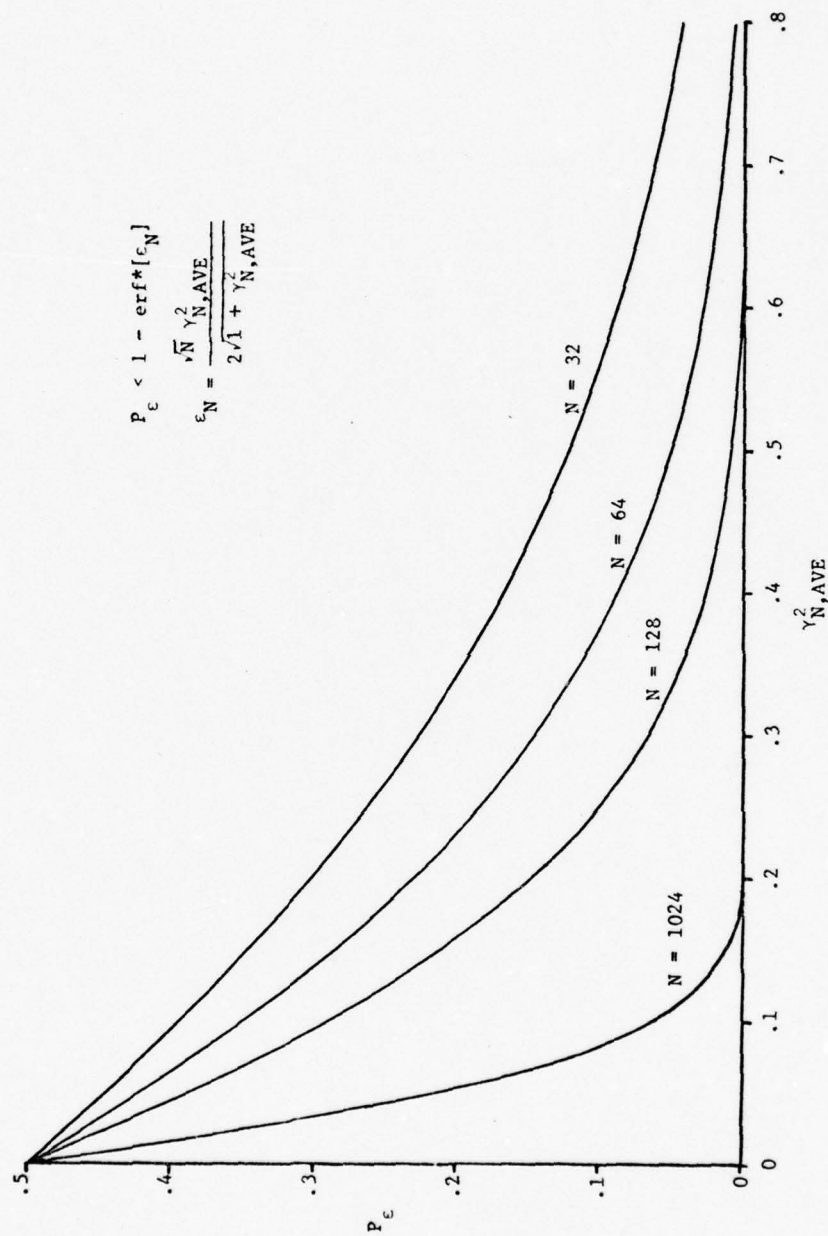
Figure 11. Maximum Probability of Error

67

$$P_\epsilon = 1 - \text{erf}^*[\epsilon_N]$$

$$\epsilon_N = \frac{\sqrt{N}\ \gamma^2_{N,AVE}}{2\sqrt{1 + \gamma^2_{N,AVE}}}$$

N = 32
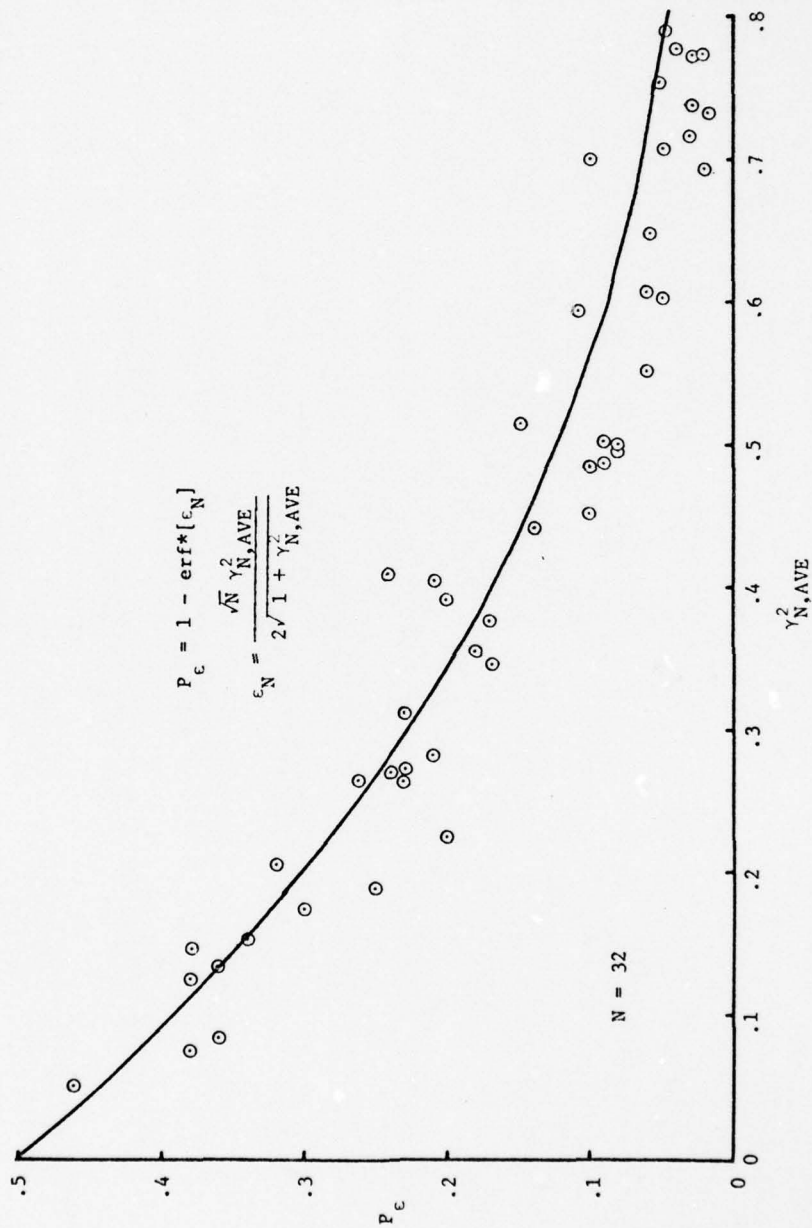
Figure 12. Experimental Probability of Error for N=32 and Uncorrelated Comparison Sets

68

$$P_\varepsilon = 1 - erf*[\varepsilon_N]$$

$$\varepsilon_N = \frac{\sqrt{N}\ \gamma^2_{N,AVE}}{2\sqrt{1 + \gamma^2_{N,AVE}}}$$
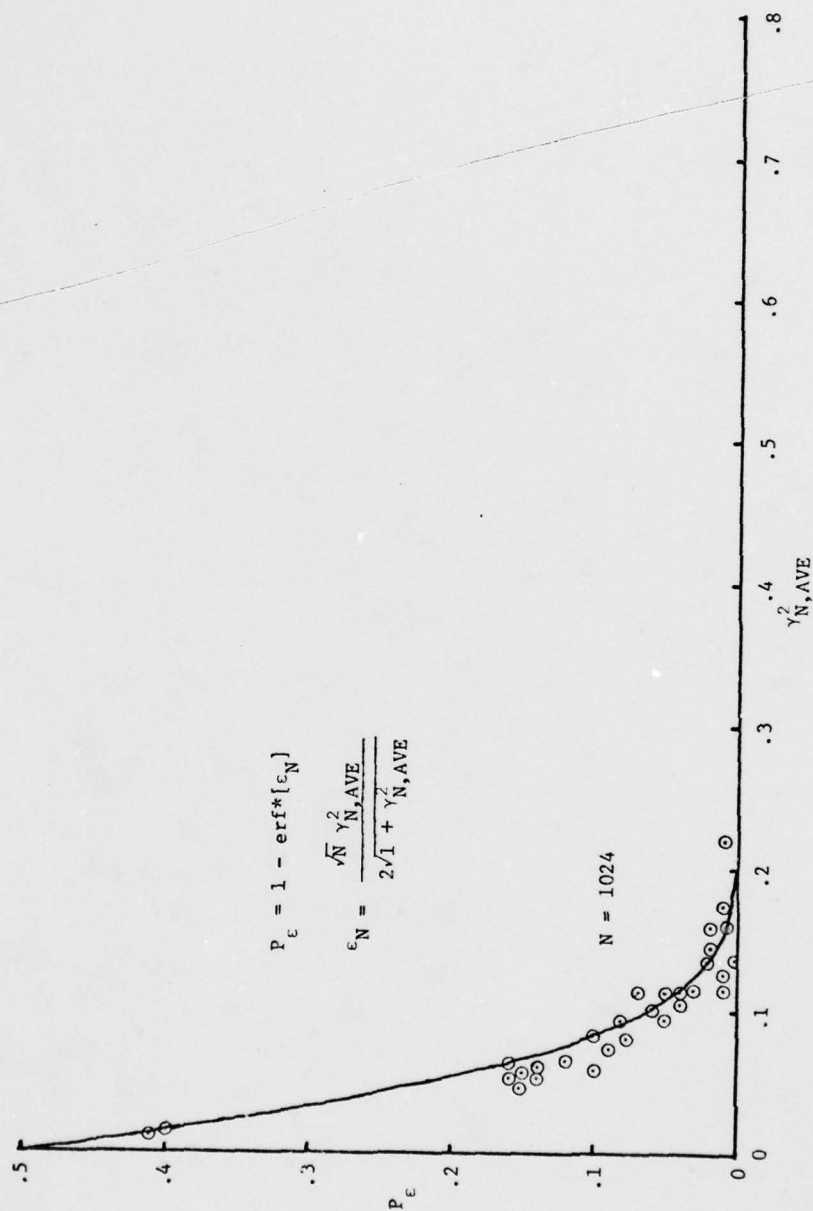
$$N = 1024$$

Figure 13. Experimental Probability of Error for N=1024 and Uncorrelated Comparison Sets

69
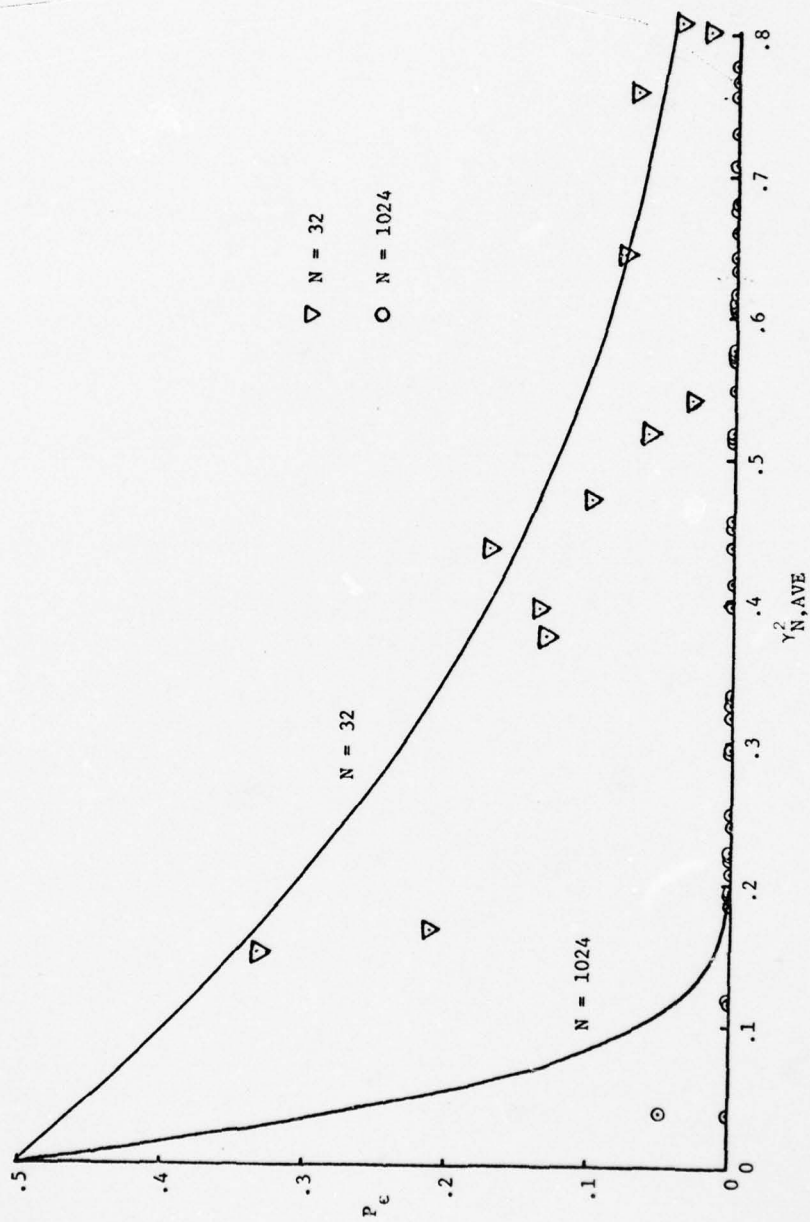
Figure 14. Experimental Probability of Error for N=32 and N=1024 with Correlated Comparison Sets

In each case the reference index set spacing was selected in

such a way as to make $d_I^2(dy,dx)$ independent of $d_I^2(0,0)$ for the points

plotted. Figure 14 illustrates the decrease in probability of error

that occurs when $d_I^2(dy,dx)$ and $d_I^2(0,0)$ are correlated. The reference

set in this case was a contiguous block of pixels. It is clear from

these results that correlation between correctly registered and

misregistered distance function values may be exploited to reduce the

probability of error below the upper bound established by (3.57).

3.12   Summary

Since $\gamma_{N,AVE}^2(dy,dx)$ depends on the sum of $\sigma_{n,REF}^2$ and $\sigma_{n,DATA}^2$,

the probability of error can be reduced by reducing either of these two

quantities. A trade-off will generally be necessary with respect to

reductions in $\sigma_{n,DATA}^2$ for line scanning sensors. Inserting a low pass

filter into the sensor output will reduce the noise variance after

digitization but will also increase the correlation time for the noise

and decrease the bandwidth of the sensor. The effect of reduced sensor

bandwidth is to decrease the magnitude of $m_i(0,dx)$ from what it would

be without a low-pass filter. While it is clearly necessary to provide

some low-pass filtering of the sensor output to suppress aliasing, the

decision on whether to reduce the bandwidth of the sensor any further

must be based on the anticipated scene content and sensor noise.

71

Filtering can also be performed after sampling and/or digitization. In Section 4.2 we will discuss a promising nonlinear technique.

The reduction of $\sigma^2_{n,REF}$ through processing of the raw data images to form a low-noise reference image must be viewed as a strong candidate. The objective is to register the sequence of incoming data images and filter the time series presented at each pixel location to form an estimate of the intensity at the corresponding point in the underlying image. Filtering on a frame-to-frame basis preserves the resolution of the raw sensor data and reduces the noise level of the reference image by averaging the noise over a number of frames. It also incorporates any changes in the underlying image into the reference image, though necessarily with some delay. Chapter 5 will deal with this subject in greater detail.

The only remaining variable to affect $\gamma^2_{N,AVE}$ is $m^2_{AVE}(dy,dx)$, the square of the average gradient of the image in the neighborhood of the set of pixels which comprise the comparison set. It is clearly possible to select the set of pixels to be included in the comparison set based on the gradient. In Section 4.3 we will discuss an approach to this selection process based on a new gradient magnitude estimation algorithm.

An alternate approach to maximizing $P_c(dy,dx)$ is to find an appropriate weighting matrix A which weights the "better" components of Ic more heavily. Section 4.1 contains the development and discussion of this approach.

72

Chapter 4

SIGNAL-TO-NOISE ENHANCEMENT TECHNIQUES

In this chapter three new techniques are developed to reduce
the probability that the tracking algorithm will make a registration
error. Each technique is demonstrated and two of them are included in
the integrated tracking algorithm which is developed and evaluated in
Chapter 6.

4.1   Nonuniformly Weighted Norm

In the previous chapter, we dealt with the case of uniform
weights for each component of Ic. In this section, we will derive an
expression for the weights a_{ii} which maximizes the lower bound on
$P_C(dy,dx)$ for a given set of $\{m_i(dy,dx)\}$ and noise variance σ_n^2.

Let

$$Y(dy,dx) \triangleq \frac{d_A^2(0,0) - d_A^2(dy,dx)}{\sigma_n^2} \qquad (4.1)$$

and normalize $Y(dy,dx)$ to zero mean and unit variance with the
transformation

73

$$Z(dy,dx) = \frac{Y(dy,dx) - E\left[Y(dy,dx)\right]}{\sqrt{var\ Y(dy,dx)}} \qquad (4.2)$$

For large N, $Z(dy,dx)$ converges in distribution to a unit normal random variable [16].

$$P_c(dy,dx) = P\left[Y(dy,dx) < 0\right]$$

$$= P\left[Z(dy,dx) < \frac{-E\left[Y(dy,dx)\right]}{\sqrt{var\ Y(dy,dx)}}\right]$$

$$= erf^*\left[\frac{-E\left[Y(dy,dx)\right]}{\sqrt{var\ Y(dy,dx)}}\right] \qquad (4.3)$$

where

74

$$E[Y(dy,dx)] = \frac{1}{\sigma_n^2}\left(E[d_A^2(0,0)] - E[d_A^2(dy,dx)]\right)$$

$$= \frac{1}{\sigma_n^2}\left\{\sum_{i=1}^{N} a_{ii}\left[E\left([n_r(0,0,i) - n_d(0,0,i)]^2\right)\right.\right.$$

$$\left.\left. -E\left([n_r(0,0,i) - m_i(dy,dx) - n_d(dy,dx,i)]^2\right)\right]\right\}$$

$$= -\frac{1}{\sigma_n^2}\sum_{i=1}^{N} a_{ii} m_i^2(dy,dx) \qquad (4.4)$$

Since $E[Y(dy,dx)]$ is negative, the lower bound on $P_c(dy,dx)$ represents the case in which $var[Y(dy,dx)]$ takes on its maximum value.

$$var[Y(dy,dx)] < \frac{1}{\sigma_n^4}\left[var[d_A^2(0,0)] + var[d_A^2(dy,dx)]\right] \qquad (4.5)$$

75

$$\text{var}[d_A^2(0,0)] = \text{var}\left[\sum_{i=1}^{N} a_{ii} \, Ic(0,0,i)^2\right]$$

$$= \sum_{i=1}^{N} a_{ii}^2 \, \text{var}[Ic(0,0,i)^2]$$

$$= 2 \, \sigma_n^4 \sum_{i=1}^{N} a_{ii}^2 \qquad\qquad (4.6)$$

$$\text{var}[d_A^2(dy,dx)] = \text{var}\left[\sum_{i=1}^{N} a_{ii} \, Ic(dy,dx,i)^2\right]$$

$$= \sum_{i=1}^{N} a_{ii}^2 \, \text{var}[Ic(dy,dx,i)^2]$$

$$= 2 \, \sigma_n^2 \sum_{i=1}^{N} a_{ii}^2 \, [2m_i^2(dy,dx) + \sigma_n^2] \qquad (4.7)$$

Substituting (4.6) and (4.7) into (4.5) yields

$$\text{var}[Y(dy,dx)] < \frac{4}{2\sigma_n} \sum_{i=1}^{N} a_{ii}^2 \, [m_i^2(dy,dx) + \sigma_n^2] \qquad (4.8)$$

Again substituting (4.4) and (4.8) into (4.3), we get the desired expression for the lower bound of $P_c(dy,dx)$.

76

$$P_c(dy,dx) \geq \text{erf}^* \left[\frac{\dfrac{1}{\sigma_n} \displaystyle\sum_{i=1}^{N} a_{ii} m_i^2(dy,dx)}{2\sqrt{\displaystyle\sum_{i=1}^{N} a_{ii}^2 [m_i^2(dy,dx) + \sigma_n^2]}} \right] \tag{4.9}$$

Since we are interested in finding the matrix coefficients a_{ii} which maximize $P_c^*(dy,dx)$, we observe that $\text{erf}(.)$ is a monotonic increasing function, and it is only necessary to maximize the argument in order to maximize the value of the function. Let

$$\varepsilon_N' = \frac{\displaystyle\sum_{i=1}^{N} a_{ii} m_i^2(dy,dx)}{2\sigma_n \sqrt{\displaystyle\sum_{i=1}^{N} a_{ii}^2 m_i^2(dy,dx) + \sigma_n^2}} \tag{4.10}$$

We take the partial derivative of $\varepsilon_N'(dy,dx)$ with respect to each non-zero element of A, and set the resulting expressions to zero.

77

$$\frac{\partial \varepsilon'_N(dy,dx)}{\partial a_{kk}} = 0 \quad ; \; k = 1,2,\ldots,N$$

$$= \frac{m_k^2}{2\sigma_n \left[\sum_{i=1}^{N} a_{ii}^2 (m_i^2 + \sigma_n^2)\right]^{\frac{1}{2}}}$$

$$- \frac{a_{kk} (m_k^2 + \sigma_n^2) \sum_{i=1}^{N} a_{ii}^2 m_i^2}{2\sigma_n \left[\sum_{i=1}^{N} a_{ii}^2 (m_i^2 + \sigma_n^2)\right]^{\frac{3}{2}}} \qquad (4.11)$$

Collecting terms containing a_{kk} and m_k yields

$$\frac{a_{kk} [m_k^2(dy,dx) + \sigma_n^2]}{m_k^2(dy,dx)} = \frac{\sum_{i=1}^{N} [m_i^2(dy,dx) + \sigma_n^2]}{\sum_{i=1}^{N} a_{ii}^2 m_i^2(dy,dx)} \qquad (4.12)$$

78

for each k, k=1,2,...,N. We now have N simultaneous equations, all of which have the same term on the right-hand side. Since the term on the right-hand side of (4.12) does not depend on k, it is an arbitrary constant. We set this constant to one,

$$\frac{\sum_{i=1}^{N} a_{ii}^2 [m_i^2(dy,dx) + \sigma_n^2]}{\sum_{i=1}^{N} a_{ii}^2 m_i^2(dy,dx)} = 1 \qquad (4.13)$$

then

$$a_{kk}(dy,dx) = \frac{m_k^2(dy,dx)}{m_k^2(dy,dx) + \sigma_n^2}$$

$$= \frac{\gamma_k^2(dy,dx)}{1 + \gamma_k^2(dy,dx)} \qquad (4.14)$$

Taking the second derivative of (4.10)

$$\frac{\partial^2 \varepsilon'_N}{\partial a_{kk}^2} = \frac{1}{\left[\sum_{i=1}^{N} a_{ii}^2 (m_i^2 + \sigma_n^2)\right]^{\frac{1}{2}}} - \frac{2a_{kk}\, m_k^2 (m_k^2 + \sigma_n^2)}{\sum_{i=1}^{N} a_{ii}^2 (m_i^2 + \sigma_n^2)}$$

$$-\frac{(m_k^2 + \sigma_n^2)\sum_{i=1}^{N} a_{ii}^2 m_i^2}{\sum_{i=1}^{N} a_{ii}^2 (m_i^2 + \sigma_n^2)} + \frac{3a_{kk}^2 \left[(m_k^2 + \sigma_n^2)\right]^2 \sum_{i=1}^{N} a_{ii}^2 m_i^2}{\left[\sum_{i=1}^{N} a_{ii}^2 (m_i^2 + \sigma_n^2)\right]^2}$$

$$(4.15)$$

substituting

$$\frac{\sum_{i=1}^{N} a_{ii}^2 (m_i^2 + \sigma_n^2)}{\sum_{i=1}^{N} a_{ii}^2 m_i^2} = 1$$

and

$$a_{kk} = \frac{m_k^2}{m_k^2 + \sigma_n^2}$$

from (4.13) and (4.14) yields

80

$$\frac{\partial^2 \mathcal{E}'_N}{\partial^2 a_{kk}} = \frac{m_k^2 + \sigma_n^2}{\left[\displaystyle\sum_{i=1}^{N} a_{ii}^2 (m_i^2 + \sigma_n^2)\right]^{\frac{1}{2}}} \left\{ \frac{a_{kk}^2 m_k^2}{\displaystyle\sum_{i=1}^{N} a_{ii}^2 m_i^2} - 1 \right\} < 0 \qquad (4.16)$$

Thus (4.14) maximizes $\mathcal{E}'_N$.

For any practical application of the weights derived here, their dependence on (dy, dx) must be removed. For any trial registration attempted, the actual translation relative to correct registration is unknown, and thus, the appropriate set of weights is unknown. Several approaches are possible.

If the sensor response time is long compared to τ_x, the reciprocal of the sampling rate, then it seems reasonable to assume that $m^2(0, k)$ will be consistently smaller than $m^2(1, 0)$ for small values of k (Figure 8 shows evidence of this fact). Under this condition it might be desirable to use the weights associated with minimizing the probability of error in a direction parallel to the scan direction.

If the sensor resolution is the same in both the horizontal and vertical directions it is possible to argue that the direction of the gradient vector is uniformly distributed on $[0, 2\pi)$. Assume that

81

$$m_k(dy,dx) = \nabla Ic(0,0,k) \cdot \hat{1}_{dy,dx} \sqrt{dy^2 + dx^2} \tag{4.17}$$

then

$$a_{kk}(r,\psi) = \frac{r^2 \left| \nabla Ic(0,0,k) \right|^2 \cos^2 \psi}{\sigma_n^2 + r^2 \left| \nabla Ic(0,0,k) \right|^2 \cos^2 \psi} \tag{4.18}$$

where

$$r = \sqrt{dy^2 + dx^2} \tag{4.19}$$

and ψ is the angle from $\hat{1}_r$ to the gradient vector. Now we take the expected value of $a_{kk}(r,\psi)$ over the interval $[0,2\pi)$

$$E_\psi[a_{kk}(r,\psi)] = a_{kk}(r) = \frac{1}{2\pi} \int_0^{2\pi} \frac{r^2 \left| \nabla Ic(0,0,k) \right|^2 \cos^2 \psi}{\sigma_n^2 + r^2 \left| \nabla Ic(0,0,k) \right|^2 \cos^2 \psi} \, d\psi$$

$$= 1 - \frac{1}{\sqrt{1 + \dfrac{r^2 \left| \nabla Ic(0,0,k) \right|^2}{\sigma_n^2}}} \tag{4.20}$$

82
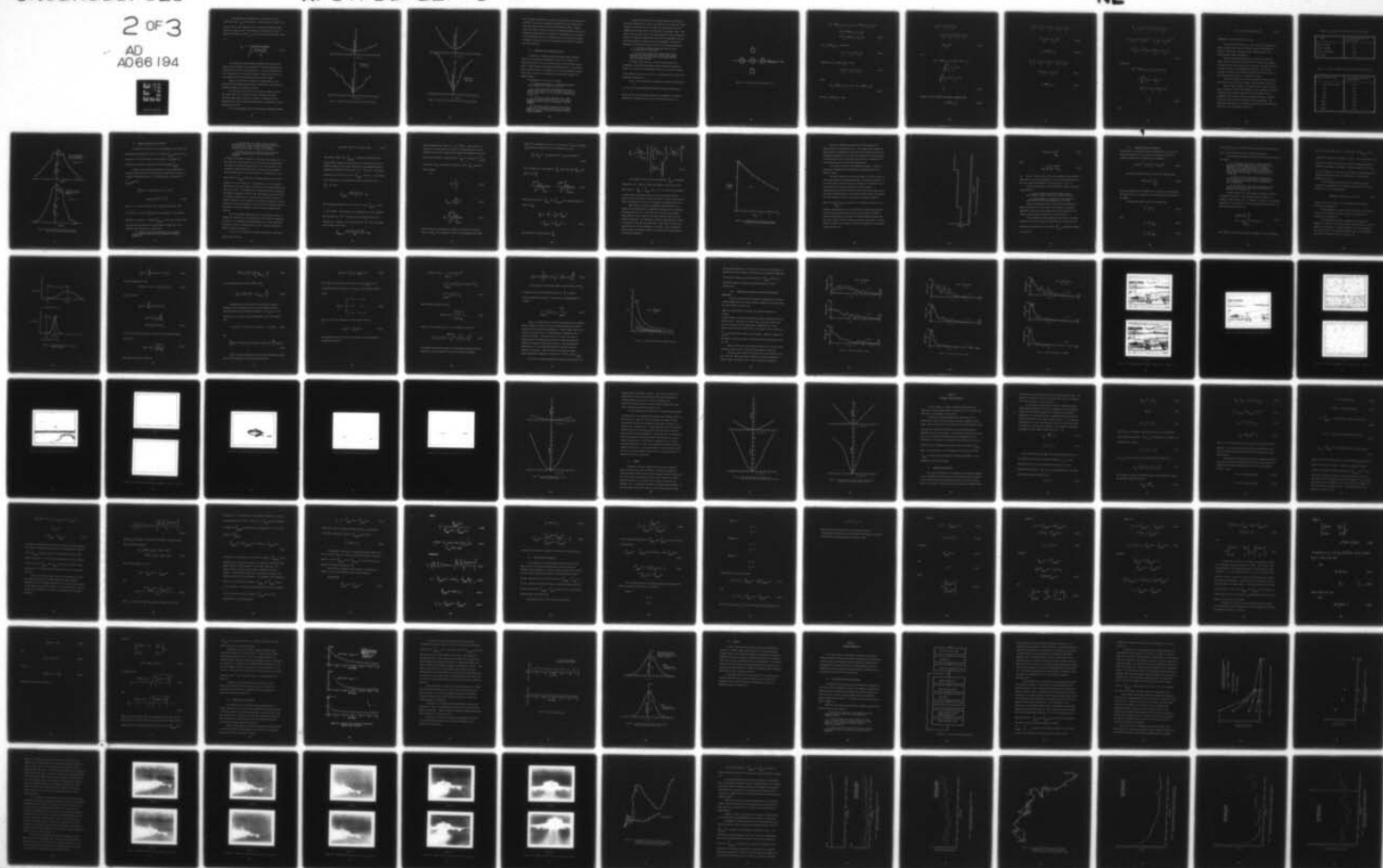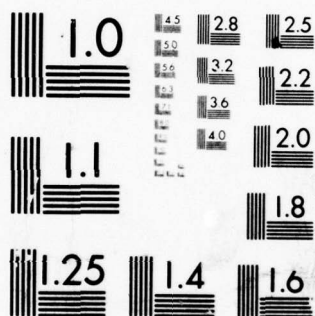
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

If a probability distribution for r was known, then the

expected value of a$_{kk}$ could be found. In the absence of a model for r,

we note that the most likely errors are those associated with small

values of r (zero or one if the tracker is working well) so that it

would seem prudent to select a$_{kk}$ to avoid the most likely errors. With

r = 1

$$a_{kk} = 1 - \frac{1}{\sqrt{1 + \dfrac{\left|\nabla Ic(0,0,k)\right|^2}{\sigma_n^2}}} \tag{4.21}$$

The benefits of using nonuniform weights when computing the

distance function depend on the presence of both relatively good and

relatively bad points in the reference set. If all of the pixels in

the reference set are of the same "quality", then they will have the

same weights, which is equivalent to having uniform weights.

Figure 15 and Figure 16 illustrate the improvement in the

normalized cross-distance function that is obtained by the use of the

nonuniform weights as specified by (4.21).

The reference set was the 32 by 32 block of adjacent pixels

with upper right corner located at row 43 and column 43 of the

reference image shown in Figure 63. Figure 15 and Figure 16 are

normalized so that the distance function value corresponding to correct

registration is 1.00.

It is of questionable value to utilize both nonuniform weights

83
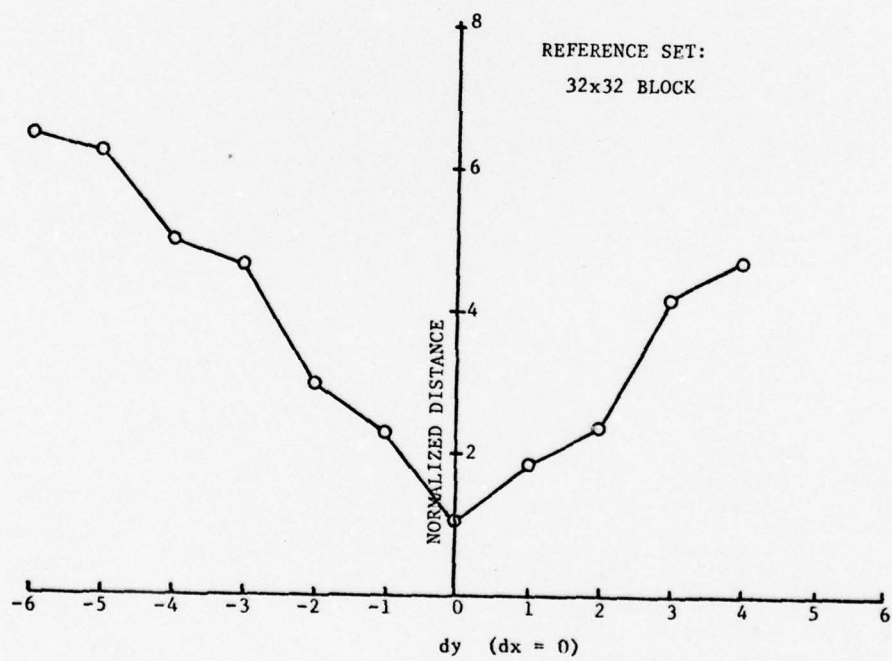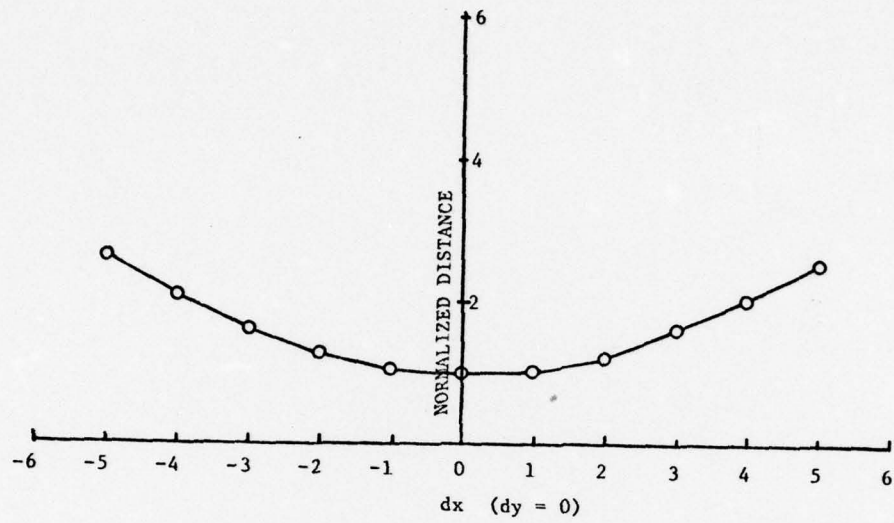
Figure 15. Normalized Cross-Distance Function Using Uniform Weights
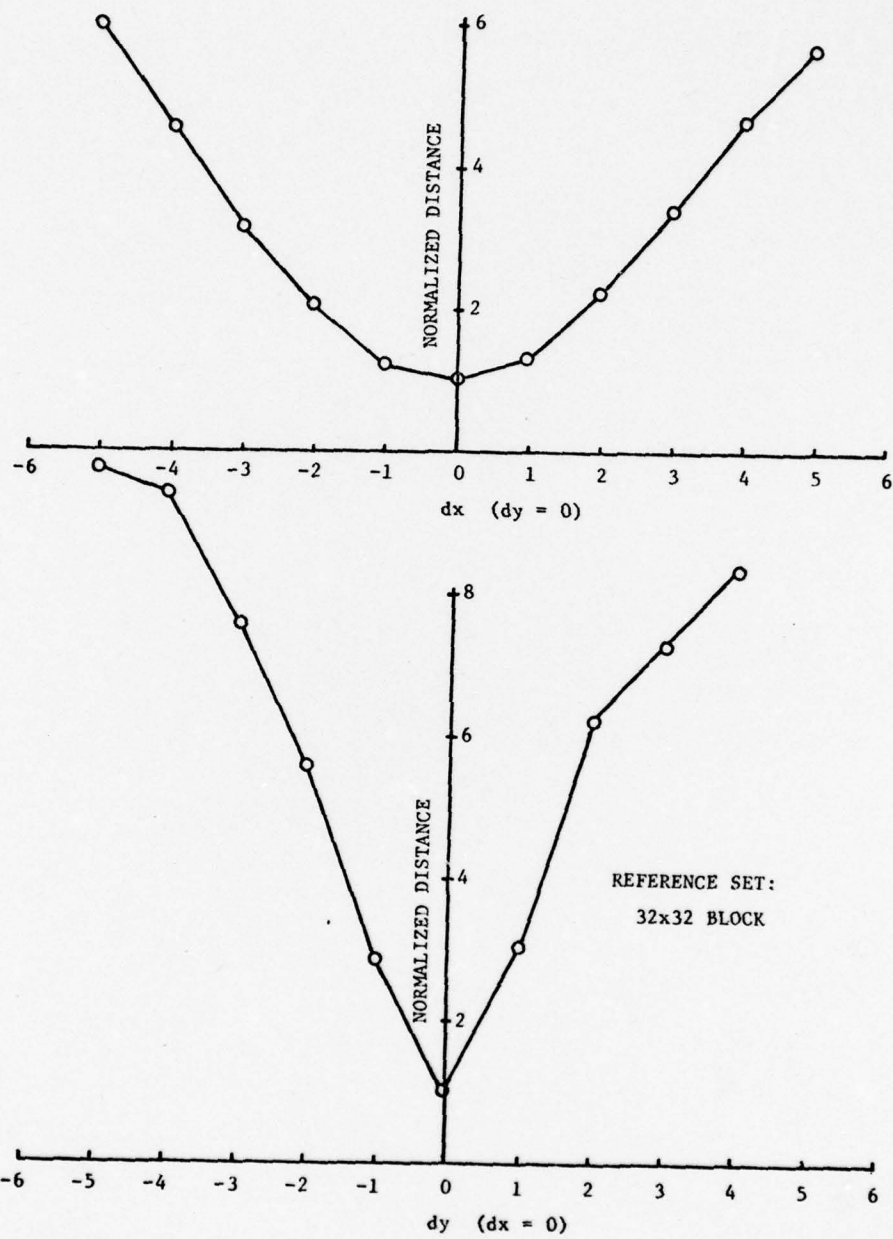
Figure 16. Normalized Cross-Distance Function Using Optimized Weights

and an adaptive reference set selection process since the reference set selection process will presumably incorporate into the reference set only those pixels which would be heavily weighted anyway. Another caveat that must be placed on the use of nonuniform weights is that the assumptions which were made to eliminate the dependence of the weights on the translation may not be appropriate for all sensors and classes of imagery, and may, in fact, be invalidated by an adaptive reference-selection algorithm.

4.2   Nonlinear Peak Elimination Filter

In addition to frequency-response-shaping filters employed ahead of the sampling and digitization steps of the tracker, the opportunity exists to filter the sensor output in the discrete domain prior to performing the similarity detection operation. The following adhoc nonlinear filter is an example of an easily mechanized algorithm which shows a potential for reducing the random noise component in sampled imagery.

The algorithm to be used is as follows:

1) Examine each pixel in an image sequentially by rows starting with the upper left corner.

2) If the pixel under examination does not have a neighbor above, below, on the left and on the right, go on to the next pixel (i.e. if it lies on an edge, don't process it).

3) If the pixel under examination has a value greater than the maximum value of its four nearest neighbors, replace it with the maximum of the four neighbors.

4) If the pixel under examination has a value smaller that the minimum value of its four nearest neighbors, replace it with the minimum of the four nearest neighbors.

86

In order to understand the motivation behind this operation, just look around and try to find a spot that is at the same time either brighter (an intensity peak) or darker (an intensity pit) than its neighbors and so small that it does not have a detectable shape. This spot, if you can find one, is the visual analog of a single resolution element in a digitized image which satisfies the requirement of being detectably brighter or darker than its surroundings. The general difficulty of finding such points leads to the following questions:

1) How often do single pixel peaks and pits occur in an image made from pure noise?

2) Is there any benefit to removing single pixel peaks and pits from images which contain both signals and noise, and if so, how big is the benefit, and how can it be characterized?

To answer these questions, we will first show that for independent, identically distributed random variables arranged and labled as in Figure 17, the probability that x_1 is either the largest or the smallest of the set of five is .4, regardless of the form of the probability distribution.

Let F_{x_i} be the cumulative distribution function associated with x_i, let f_{x_i} be the corresponding probability density function, and assume that all of the density functions are continuous. We wish to establish the probability that x_1 is a local extreme point, i.e.
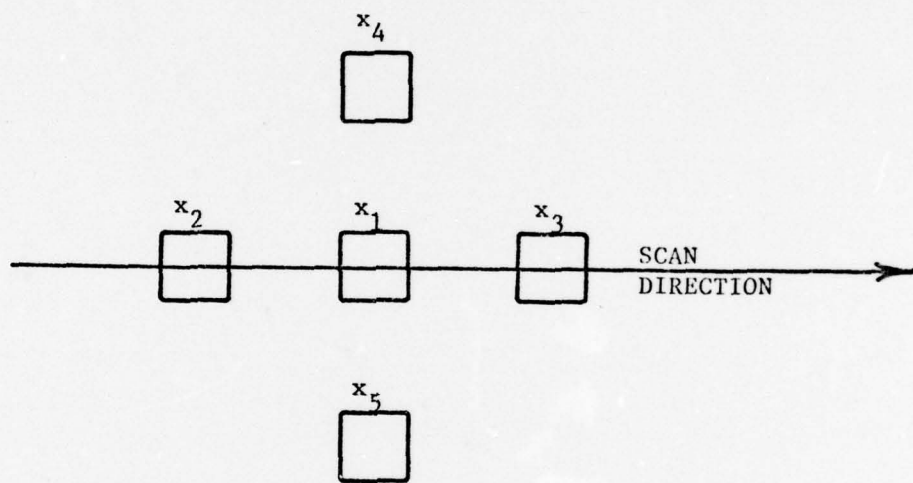
Figure 17.   The Four Nearest Neighbors of x_1

$$P[x_1 > MAX\{x_2, x_3, x_4, x_5\} \text{ or } x_1 < MIN\{x_2, x_3, x_4, x_5\}]$$

$$= P[x_1 > MAX\{x_2, x_3, x_4, x_5\}]$$

$$+ P[x_1 < MIN\{x_2, x_3, x_4, x_5\}] \tag{4.22}$$

Let $y_1 = MAX\{x_2, x_3\}$, then [26]

$$F_{y_1}(y_1) = F_{x_2}(y_1) F_{x_3}(y_1) \tag{4.23}$$

similarly, let $y_2 = MAX\{x_4, x_5\}$, then

$$F_{y_2}(y_2) = F_{x_4}(y_2) F_{x_5}(y_2) \tag{4.24}$$

Since

$$P[x_1 > MAX\{x_2, x_3, x_4, x_5\}] = P[x_1 > MAX\{y_1, y_2\}]$$

$$\tag{4.25}$$

we let $y_3 = MAX\{y_1, y_2\}$, then

89

$$F_{y_3}(y_3) = F_{y_1}(y_3) F_{y_2}(y_3)$$

$$= F_{x_2}(y_3) F_{x_3}(y_3) F_{x_4}(y_3) F_{x_5}(y_3)$$

$$= [F_{x_1}(y_3)]^4 \tag{4.26}$$

and

$$P[x_1 > MAX\{x_2, x_3, x_4, x_5\}] = P[x_1 > y_3]$$

$$= \int_{-\infty}^{\infty} f_{x_1} F_{y_3}(\lambda) \, d\lambda$$

$$= \int_{-\infty}^{\infty} [F_{x_1}(\lambda)]^4 F'_{x_1}(\lambda) \, d\lambda$$

$$= \frac{1}{5} \tag{4.27}$$

Through a similar, though more tedious, argument with

$$z_1 = MIN\{x_2, x_3\} \tag{4.28}$$

90

$$F_{z_1}(z_1) = F_{x_2}(z_1) + F_{x_3}(z_1) - F_{x_2}(z_1) F_{x_3}(z_1)$$

$$= 2 F_{x_1}(z_1) - [F_{x_1}(z_1)]^2 \tag{4.29}$$

$$z_2 = MIN\{x_4, x_5\} \tag{4.30}$$

$$F_{z_2}(z_2) = F_{x_4}(z_2) + F_{x_5}(z_2) - F_{x_4}(z_2) F_{x_5}(z_2)$$

$$= 2 F_{x_1}(z_2) - [F_{x_1}(z_2)]^2 \tag{4.31}$$

$$z_3 = MIN\{z_1, z_2\} \tag{4.32}$$

91

$$F_{z_3} = F_{z_1}(z_3) + F_{z_2}(z_3) - F_{z_1}(z_3) F_{z_2}(z_3)$$

$$= 4 F_{x_1}(z_3) - 2 [F_{x_1}(z_3)]^2 - 2 F_{x_1}(z_3) - [F_{x_1}(z_3)]^2$$

$$= -F_{x_1}^4(z_3) + 4 F_{x_1}^3(z_3) - 6 F_{x_1}^2(z_3) + 4 F_{x_1}(z_3)$$

<div align="right">(4.33)</div>

we find that

$$P[x_1 < MIN\{x_2, x_3, x_4, x_5\}] = P[x_1 < z_3]$$

$$= 1 + \int_{-\infty}^{\infty} [F_{x_1}^4(\lambda) - 4 F_{x_1}^3(\lambda)$$

$$+ 6 F_{x_1}^2(\lambda) - 4 F_{x_1}(\lambda)] F_{x_1}'(\lambda) \, d\lambda$$

$$= \frac{1}{5}$$

<div align="right">(4.34)</div>

thus

$$P[\ x_1 \text{ is an extreme point}] = \frac{2}{5} \qquad (4.35)$$

independent of the distribution of the x_i.

How does this 40% figure for a pure noise image compare with natural imagery after it has been sensed, sampled, and digitized? Figure 56, Figure 57, and Figure 58 show samples from the image sequences CARS, TREES, and AIRPLANE.

Table III lists the percentage of pixels in each of four images that are local extrema, and Table IV illustrates the change in the percentage of local extreme points as noise of increasing variance is added to frame 1 from image sequence AIRPLANE. As the noise variance is increased, the fraction of pixels that are local peaks or pits approaches the limit of .4 predicted by theory.

Figure 18 illustrates the noise distribution before and after application of the filter for an image containing pure random noise.

Based on these results, the non-linear peak elimination filter shows considerable promise for reducing the noise component of imagery in regions of low contrast (regions with a high degree of local randomness) while leaving relatively unchanged the signal component (persistent local intensity gradients) of the same imagery. The performance improvement of a tracking system using the peak elimination prefilter will be documented in Chapter 6.

Table III. Extreme point statistics for real data images

Image	% of total pixels that are peaks or pits
Frame 1 of CARS	15.5
Frame 1 of TREES	20.7
Frame 1 of AIRPLANE	24.0
Pure noise	39.4

Table IV. Effects of additive noise on extreme point statistics

Variance of additive noise	% of total pixels that are peaks or pits
0.00 (original image)	24.0
.04	26.6
.25	30.2
1.00	34.4
4.00	36.6
16.00	38.9
25.00	38.8
100.00	39.5
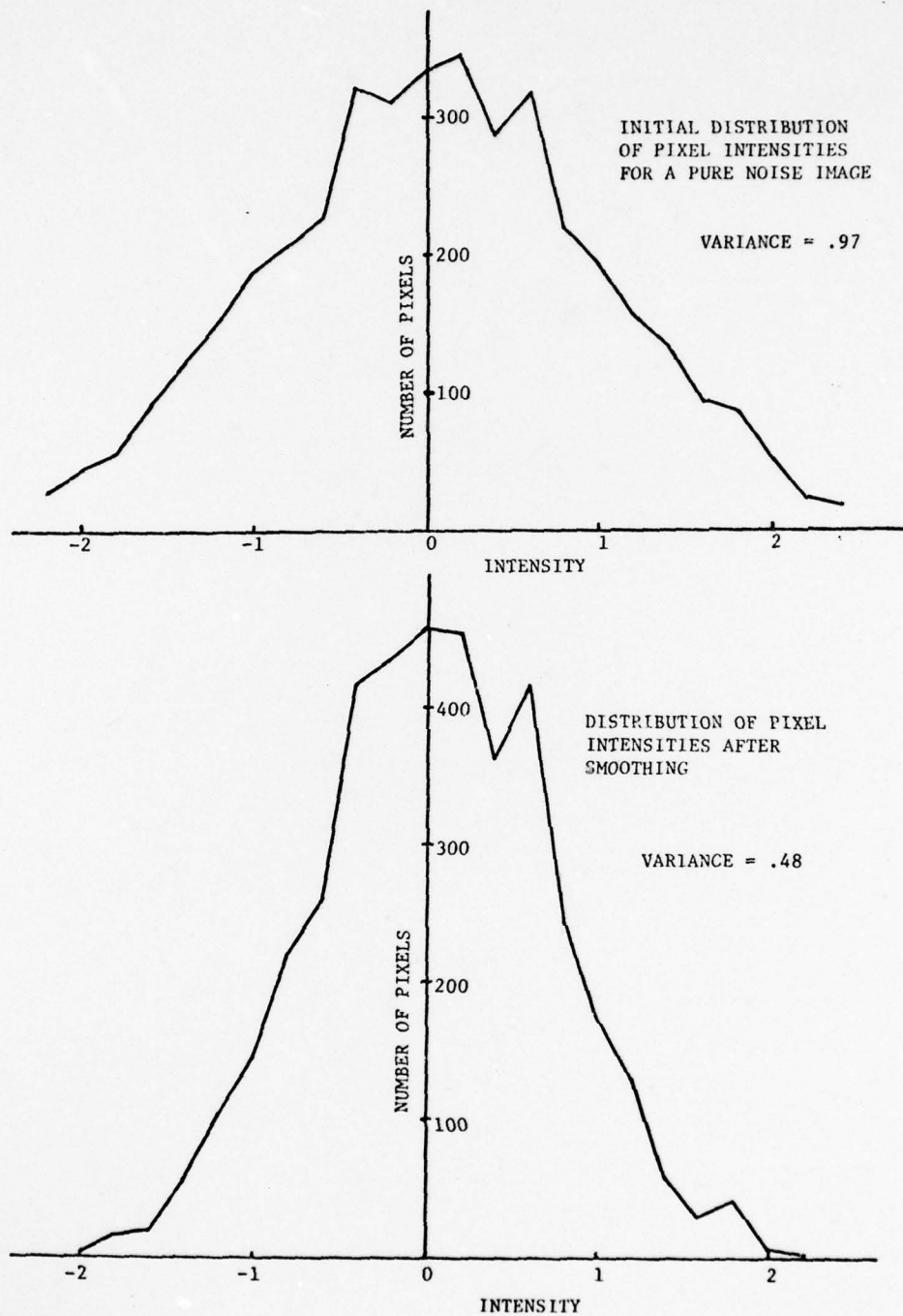
Figure 18.   Noise Distribution Before and After Smoothing
with the Non-Linear Peak Elimination Filter

95

4.3   Adaptive Reference Set Selection

In Section 3.12 we saw that it is advantageous (with respect to minimizing the probability of error) to maximize $m_{N,AVE}^2$ (dy,dx) for any particular N. In this section, we will develop a technique for adaptively selecting the reference set which maximizes $m_{N,AVE}^2$.

Consider an arbitrary pixel Ic(0,0,k) in the comparison set. The facet model assumes that the gradient is approximately constant in a region surrounding any point. The contribution of ∇Ic(0,0,k) to $m_{N,AVE}^2$ (dy,dx) is

$$m_k^2 (dy,dx) = [\nabla Ic(0,0,k) \cdot (dy\ \hat{1}_y + dx\ \hat{1}_x)]^2$$

$$= \left| \nabla Ic(0,0,k) \right|^2 r^2 \cos^2 \Psi \qquad (4.36)$$

where Ψ is the angle between $(dy\hat{1}_y + dx\hat{1}_x)$ and ∇Ic(0,0,k), and $r^2 = dy^2 + dx^2$. We see that m_k^2 (dy,dx) is proportional to the gradient magnitude at Ic(0,0,k). To maximize $m_{N,AVE}^2$, select the N pixels in the reference image which have the largest gradient magnitudes. This selection can be performed in three steps:

1) Calculate the gradient magnitude for each pixel in the reference image and form a histogram of gradient magnitudes.

96

2) Starting with the largest gradient magnitude and working down, find the largest threshold value such that there are at least N pixels with gradient magnitudes greater than or equal to the threshold.

3) Find N pixels in the reference image which have a gradient magnitude greater than or equal to the threshold.

A hazard exists with this approach to selecting the reference set. If the resolution of the sensor in the scan direction is substantially less than the resolution perpendicular to the scan direction, it is possible for this algorithm to select N pixels with all gradient vectors perpendicular to the scan direction. The result is a very small value for $m^2_{N,AVE}$ (0,dx) and an increased probability of errors parallel to the scan direction. The indication is that the angular resolution of the sensor after digitization should be approximately the same in each axis. This problem can be lessened somewhat by weighting the gradient component parallel to the scan direction more heavily than the component perpendicular to the scan direction when computing the gradient magnitude, or by using only the component parallel to the scan direction to form the histogram. For the three data sequences used for evaluation of tracking algorithms, neither of these strategies was required.

Since the gradient magnitudes are used in decreasing order to assemble the reference set, the possibility exists that the probability of error is not a monotone decreasing function of N. What conditions would have to exist in order for this to occur, and what precautions should be taken to prevent it?

If P_ε is not a monotonic decreasing function of N, then there exists some k such that

97

$$P_{\mathcal{E}}\,[dy, dx \mid N=k+1] > P_{\mathcal{E}}\,[dy, dx \mid N=k] \qquad (4.37)$$

For example, suppose that $\gamma^2_{50,AVE}$ = .160 when the 50 pixels with larges gradient magnitude are included in the reference set, and in addition, suppose that the next 200 pixels in the gradient magnitude histogram had signal-to-noise ratios of .07. From (3.57) the maximum probability of error for N = 50 and $\gamma^2_{N,AVE}$ = .160 is 1 - erf [.52523]*. If the next pixel to be incorporated into the reference set has γ^2_{51} = .07, then

$$\gamma^2_{51,AVE} = \frac{.160 \times 50 + .07}{51} = .158$$

and the maximum probability of error for N = 51 and $\gamma^2_{N,AVE}$ = .158 is 1 - erf [.52500]*. This increase in the probability of error indicates that the pixel with γ^2 = .07 should not be incorporated into the reference set. However, if all 200 pixels with γ^2 = .07 are included in the reference set, then

$$\gamma^2_{250,AVE} = \frac{.160 \times 50 + .07 \times 200}{250} = .088$$

98

and the probability of error is 1 - erf* [.66697]. From this we see
that there are cases where the inclusion of additional pixels in the
reference set does not automatically decrease the probability of error.
We will next establish a lower bound for γ^2_{N+1} as a function of $\gamma^2_{N,AVE}$
and N such that P_ε will always decrease so long as γ^2_{N+1} is greater
than the bound.

Let

$$\gamma^2_i = \frac{m^2_i}{\sigma^2_n} \qquad (4.38)$$

$$\gamma^2_{N,AVE} = \frac{1}{N} \sum_{i=1}^{N} \gamma^2_i \qquad (4.39)$$

$$\bar{\varepsilon}_N = \frac{\sqrt{N} \; \gamma^2_{N,AVE}}{2\sqrt{1 + \gamma^2_{N,AVE}}} \qquad (4.40)$$

where we drop the specification of (dy,dx) to simplify the notation.
Recall that $\bar{\varepsilon}_N$ is the argument of erf*(.) which establishes the upper

99

bound on the probability of error. From (3.57) we see that increasing $\bar{\mathcal{E}}_N$ decreases the upper bound on the probability of error

$$\bar{\mathcal{E}}_N < \bar{\mathcal{E}}_{N+1} \Rightarrow \quad P_{\mathcal{E}} [dy, dx \mid N=k] > P_{\mathcal{E}} [dy, dx \mid N=K+1]$$

(4.41)

We are looking for a lower bound on γ^2_{N+1} which will cause $\bar{\mathcal{E}}_{N+1}$ to be greater than $\bar{\mathcal{E}}_N$

$$\frac{\sqrt{N} \; \gamma^2_{N,AVE}}{2\sqrt{1 + \gamma^2_{N,AVE}}} < \frac{\sqrt{N+1} \; \gamma^2_{N,AVE}}{2\sqrt{1 + \gamma^2_{N+1,AVE}}}$$

(4.42)

substituting (4.39) for $\gamma^2_{N,AVE}$ and $\gamma^2_{N+1,AVE}$ and collecting terms in powers of γ_{N+1}

$$\gamma^4_{N+1} [1 + \gamma^2_N] + \gamma^2_{N+1} N[(\gamma^2_{N,AVE})^2$$

$$+ 2 \gamma^2_{N,AVE}] - N(\gamma^2_{N,AVE})^2 > 0$$

(4.43)

and solving for the lower bound on γ^2_{N+1}

100

$$\gamma^2_{N+1} > \frac{N \, \gamma^2_{N,AVE}}{1 + \gamma^2_{N,AVE}} \left\{ \left[\left(\frac{\gamma^2_{N,AVE}}{2} + 1 \right)^2 + \frac{\gamma^2_{N,AVE} + 1}{N} \right]^{\frac{1}{2}} - \left[\frac{\gamma^2_{N,AVE}}{2} + 1 \right] \right\} \tag{4.44}$$

For N greater than 10, this lower bound for γ^2_{N+1} is relatively insensitive to N. Figure 19 shows the behavior of the ratio of the lower bound for γ^2_{N+1} to $\gamma^2_{N,AVE}$ for $N > 10$. At each step, the signal-to-noise ratio for the next pixel divided by the current average signal-to-noise ratio must lie above the curve to assure that the upper bound on probability of error is a monotonic decreasing function of N.

There seems to be only one realistic situation where it is likely that the upper bound on probability of error is not a monotonic decreasing function of N. This case will occur when the reference image contains a small number of very high contrast pixels on a low contrast background. The histogram of gradient magnitudes will contain a few points in the high value bins with a large span of vacant bins separating these from the remainder of the image. This case did not arise in any of the imagery used for the evaluation of this reference set selection algorithm.
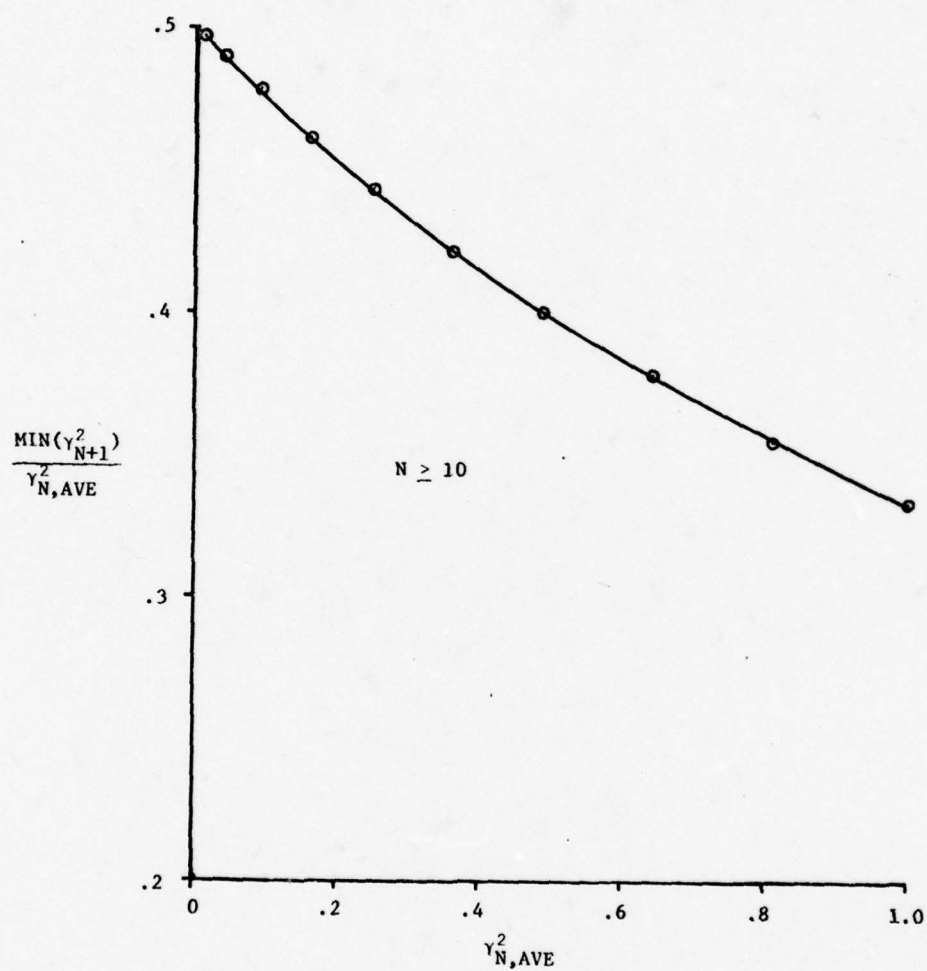
101

Figure 19. Incremental Signal-to-Noise Ratio Requirements
for Monotonic Decreasing Probability of Error

There is an additional restriction on the inclusion of a particular pixel in the reference set. For integer-valued imagery, a pixel should be considered for inclusion in the reference set only if the gradient magnitude in the region surrounding the pixel in question is sufficiently large to assure that there will be a contribution to the distance function for some trial registration.

To see this, consider a one-dimensional example. Figure 20 illustrates a hypothetical one-dimensional intensity profile for a digitized image.

The pixel labeled X can be shifted right or left by as much as four pixels without contributing anything to the distance function. If the search region is plus and minus two pixels, the inclusion of X in the reference set contributes nothing to discovering any misregistration between the reference image and the data image, even if the actual misregistration is the maximum allowable value (two pixels). The local average derivative in the region surrounding X must be greater than $\dfrac{1}{4R_{MAX} + 1}$ to contribute to the distance function at any allowable misregistration, and a more practical limit would require a local average derivative sufficiently large to produce a distance function contribution at the edge of the search region even when the data image and the reference image are perfectly registered. For this reason, we place the following constraint on the pixels to be included in the reference set:
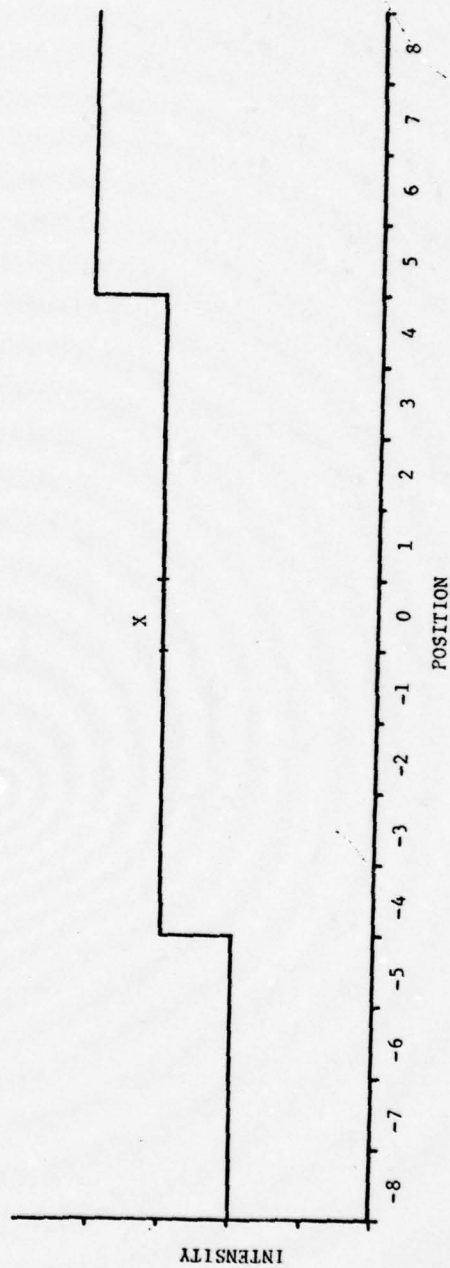
Figure 20. Hypothetical One-Dimensional Digitized Image

$$|\nabla Ir(i,j)| > \frac{1}{R_{MAX}} \qquad (4.45)$$

where

$$R_{MAX} = \sqrt{dy_{MAX} + dx_{MAX}} \qquad (4.46)$$

For   Ir(i,j) less than this limit, and a perfectly registered data

image there is no trial registration in a noise-free image which has

any error signal attributable to the inclusion of (i,j) in L.

There are two types of regions which will exhibit the desired

characteristic:

1) Local extreme points (either peaks or pits)
will exhibit a change in intensity in every direction.

2) Points which lie on edges will show a change in
intensity in directions perpendicular to the edge.

Since any significant peak or pit will be surrounded by an

edge, the approach to be taken will be to estimate the gradient

magnitude at each point in the image and use the N points with the

largest gradient magnitude values for the reference set.

In the remainder of this chapter we will develop a gradient

magnitude estimation technique for selecting the reference set and

demonstrate the potential for increasing $\gamma_{N,AVE}^2$ by adaptive reference

set selection.

4.3.1   Gradient Magnitude Estimation

The approach to estimating the gradient magnitude at a point in the
reference image will be to estimate the horizontal and vertical
components separately and combine them according to

$$\left| \nabla Ir \right|^2 = \left| \nabla Ir \cdot \hat{1}_x \right|^2 + \left| \nabla Ir \cdot \hat{1}_y \right|^2 \tag{4.47}$$

The error to avoid is that of incorrectly estimating that

$$\left| \nabla Ir(i,j) \right| > \frac{1}{R_{MAX}} \tag{4.48}$$

and hence making $Ir(i,j)$ a candidate for inclusion in the comparison
set when it will only contribute to the noise and never contribute to
the signal.

A rectangular search area will be assumed, with

$$R^2 = R_x^2 + R_y^2 \tag{4.49}$$

where

$$R_x = \eta_x \, dx_{MAX} \tag{4.50}$$

$$R_y = \eta_y \, dy_{MAX} \tag{4.51}$$

106

and η_x and η_y are the scale factors which convert pixel spacing in x

and y to horizontal and vertical angular displacements.

The following assumptions are made about the characteristics of the image:

   1) The distance over which a gradient component
persists is roughly proportional to the inverse of its
magnitude (i.e. low gradients persist for longer
distances than do higher gradients). If this was not
true, then a histogram of gradient magnitudes would
cluster away from the origin. As we shall see, this is
not the case.

   2) The occurance of single pixel extreme points
that are not due to noise phenomena is relatively rare
in a randomly selected image. Table 3 indicates that
this is reasonable.

   3) The covariance of the noise in the image can be
modeled as a zero-mean, first-order Markov process in
each axis (see Section 3.8).

In order to bound the rate at which low gradient points are

erroneously determined to have sufficient signal strength to be

included in the reference set, a restriction is placed on the allowable

performance of the estimator. For each component of the gradient, the

following criterion must be met:

$$\frac{\eta_i \left| \hat{\nabla} I r \cdot \hat{1}_i \right| - \frac{\eta_i}{R_i}}{\sigma_{\hat{\nabla} I r \cdot \hat{1}_i}} \geq \text{constant} = C \qquad (4.52)$$

where $\hat{\nabla} I r \cdot \hat{1}_i$ is the estimated gradient component in the $\hat{1}_i$ direction,

107

R_i is the search radius in the $\hat{1}_i$ direction, and $\sigma_{\widehat{\nabla Ir \cdot \hat{1}_i}}$ is the

standard deviation of the estimate of $\nabla Ir \cdot \hat{1}_i$. This restriction on the

estimator will insure that there are at least C standard deviations

between the estimate of the gradient component magnitude and the $1/R_i$

point (see Figure 21). The effect is to build an estimator which has a

variable confidence interval, but operates with a fixed maximum

probability of error.

With this approach to estimating the gradient magnitude

components, the probability of erroneously including a point in the

reference set can be approximated as

$$P[\text{ERRONEOUS INCLUSION}] = [\text{erf}^*(-C)]^2 \qquad (4.53)$$

where we assume that errors in the x-component and y-component

estimates are independent.

This probability can be made arbitrarily small by increasing C

at the cost of reducing the number of pixel locations that are

candidates for inclusion in L.

In estimating the gradient components, each component is

modeled as a constant in the region used to form the estimate. For the

purpose of developing the necessary equations, only the x-component

will be dealt with. The y-component differs only in the correlation

time used in the Markov model for the noise. Let

108
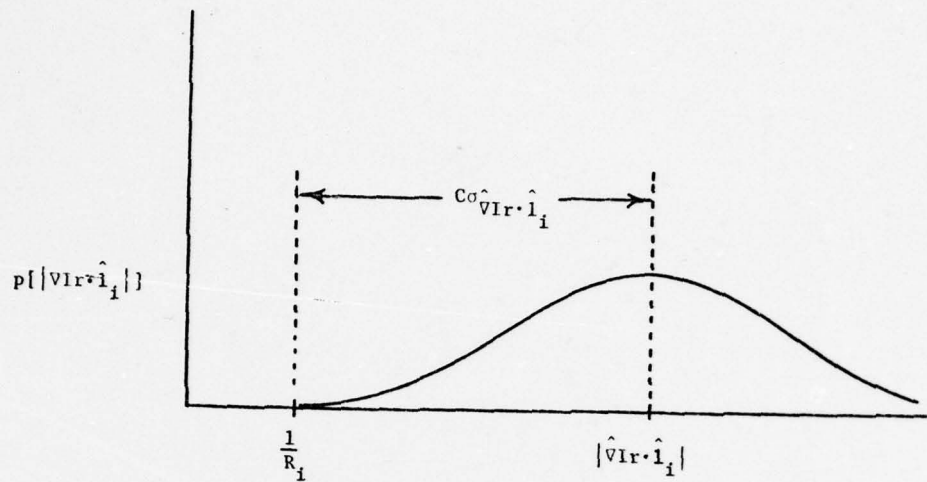
a) A LARGE VALUE OF $\nabla Ir \cdot \hat{1}_i$
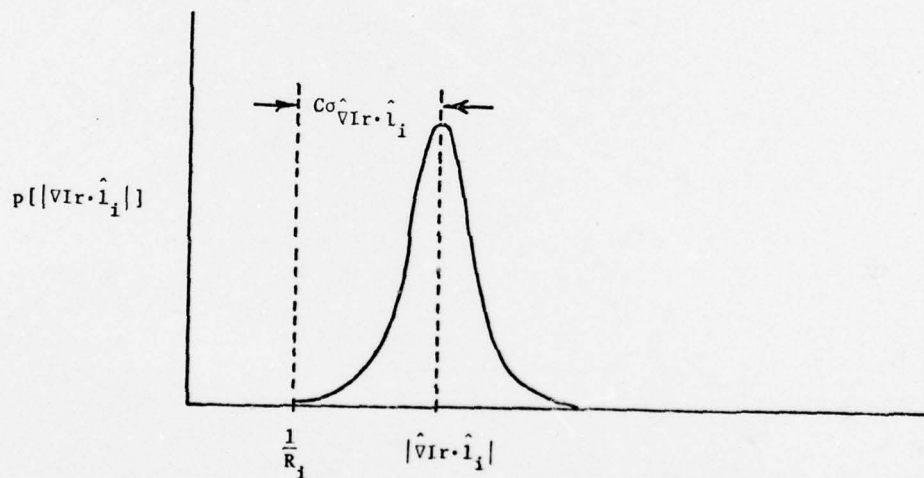
b) A SMALL VALUE OF $\nabla Ir \cdot \hat{1}_i$

Figure 21. Desired Characteristics of a Gradient
Magnitude Estimator

109

$$Sx_K(i,j) = \sum_{n=1}^{K} Ir(i,j+n) - Ir(i,j-n) \qquad (4.54)$$

With the approximation that

$$Ir(i,j+n) = Ir(i,j) + n\,\eta_x \,\nabla Ir(i,j)\cdot\hat{1}_x \qquad (4.55)$$

$Sx_k(i,j)$ becomes

$$Sx_K(i,j) = \sum_{n=1}^{K} 2n\,\eta_x \,\nabla Ir(i,j)\cdot\hat{1}_x$$

$$= 2\,\nabla Ir(i,j)\cdot\hat{1}_x \,\eta_x \sum_{i=1}^{N} n$$

$$= K(K+1)\,\eta_x \,\nabla Ir(i,j)\cdot\hat{1}_x \qquad (4.56)$$

From 4.58 we define the estimate for $\nabla Ir(i,j)\cdot\hat{1}_x$ using 2k adjacent

pixels to be

$$\hat{\nabla} Ir(i,j)\cdot\hat{1}_x = \frac{Sx_K(i,j)}{K(K+1)\,\eta_x} \qquad (4.57)$$

The required constraint reduces to

110

$$\left| \hat{\nabla} Ir(i,j) \cdot \hat{1}_x \right| \geqslant \frac{c}{\eta_x} \sigma_{\hat{\nabla} Ir \cdot \hat{1}_x} + \frac{1}{R_x} \qquad (4.58)$$

or after substituting (4.57) for $\hat{\nabla} Ir(i,j) \cdot \hat{1}_x$

$$\left| Sx_K(i,j) \right| \geqslant K(K+1) \; c \; \sigma_{\hat{\nabla} Ir \cdot \hat{1}_x} + \frac{\eta_x}{R_x} \qquad (4.59)$$

Determination of the variance of the estimate requires a knowledge of the covariance of the noise associated with the 2k adjacent pixels used in forming the estimate. Let $zx_K(i,j)$ be the vector made up of the set of 2k+1 pixels used to form the estimate, then

$$[zx_K(i,j)]^T = [Ir(i,j-K), \; Ir(i,j-K+1), \ldots, Ir(i,j+K)] \qquad (4.60)$$

and

$$E\left[\left(zx_K(i,j) - E[zx_K(i,j)] \right) \left(zx_K(i,j) - E[zx_K(i,j)] \right)^T \right] \triangleq Px_K(i,j)$$

$$(4.61)$$

Since the noise is modeled as a first-order Markov process with zero mean, the elements of the covariance matrix are

111

$$\underset{K}{Px}(i,j) = \underset{n}{\sigma}^{2} \exp(-\mu_{x}\eta_{x}|i-j|) \qquad (4.62)$$

where $\underset{n}{\sigma}^{2}$ is the variance of the error in Ir, and $\mu_{x}\eta_{x}$ is the

correlation time for the noise. Let $\underset{K}{\omega}$ be a 2k+1 element constant

vector

$$\underset{K}{\omega}(i) = \begin{cases} -1 & \text{for} \quad 1 < i < K \\ \\ 0 & \text{for} \quad i = K \\ \\ 1 & \text{for} \quad K < i < 2K \end{cases} \qquad (4.63)$$

$\underset{K}{Sx}(i,j)$ can now be expressed in a more compact notation

$$\underset{K}{Sx}(i,j) = \underset{K}{\omega}^{T} \underset{K}{zx}(i,j) \qquad (4.64)$$

An expression for the variance of the estimate is now obtained in a
straightforward manner

$$\text{var}[\hat{\nabla} Ir(i,j) \cdot \hat{1}_x] = \frac{1}{K^2(K+1)^2} E[Sx_K(i,j)^2]$$

$$= \frac{1}{K^2(K+1)^2} E[\omega_K^T \, zx_K(i,j) \, zx_K^T(i,j) \, \omega_K]$$

$$= \frac{1}{k^2(K+1)^2} \omega_K^T \, Px_K \, \omega_K \qquad (4.65)$$

The estimator now takes the form

$$\hat{\nabla} Ir(i,j) \cdot \hat{1}_x = \frac{\omega_K^T \, zx_K(i,j)}{K'(K'+1)\eta_x} \qquad (4.66)$$

subject to the constraint that K' is the smallest K such that

$$\left| \omega_K^T \, zx_K(i,j) \right| \geqslant \frac{K(K+1)\eta_x}{R_x} + C\left[\omega_K^T \, Px_K \, \omega_K \right]^{\frac{1}{2}} \qquad (4.67)$$

The gradient magnitude components are individually estimated and then combined to form an estimate of the overall gradient magnitude.

113

$$\left|\hat{\nabla}Ir(i,j)\right| = \left[\left[\hat{\nabla}Ir(i,j)\cdot\hat{1}_x\right]^2 + \left[\hat{\nabla}Ir(i,j)\cdot\hat{1}_y\right]^2\right]^{\frac{1}{2}} \tag{4.68}$$

If the noise in the reference image is uncorrelated, then Px_K is diagonal with all nonzero entries equal to σ_n^2, the global reference-image error variance. This results in simplification of (4.65) to

$$\frac{1}{K^2(K+1)^2}\,\omega_K^T\,Px_K\,\omega_K = \frac{2\sigma_n^2}{K(K+1)^2} \tag{4.69}$$

There is an upper limit on K which is determined by the minimum distance between the pixel at which the gradient magnitude is being estimated and the nearest image boundary. A smaller upper limit may be desirable in practice due to considerations of computation time or hardware complexity. In either case, if the upper limit of K is reached without satisfying the constraint, the gradient magnitude component can simply be estimated to be zero. This precludes the possibility of an increasing classification error rate near the boundary of the image. Figure 22 shows the minimum detectable gradient-magnitude component as a function of $\nabla Ir\cdot\hat{1}_i$ and $C\sigma_{n,REF}$.

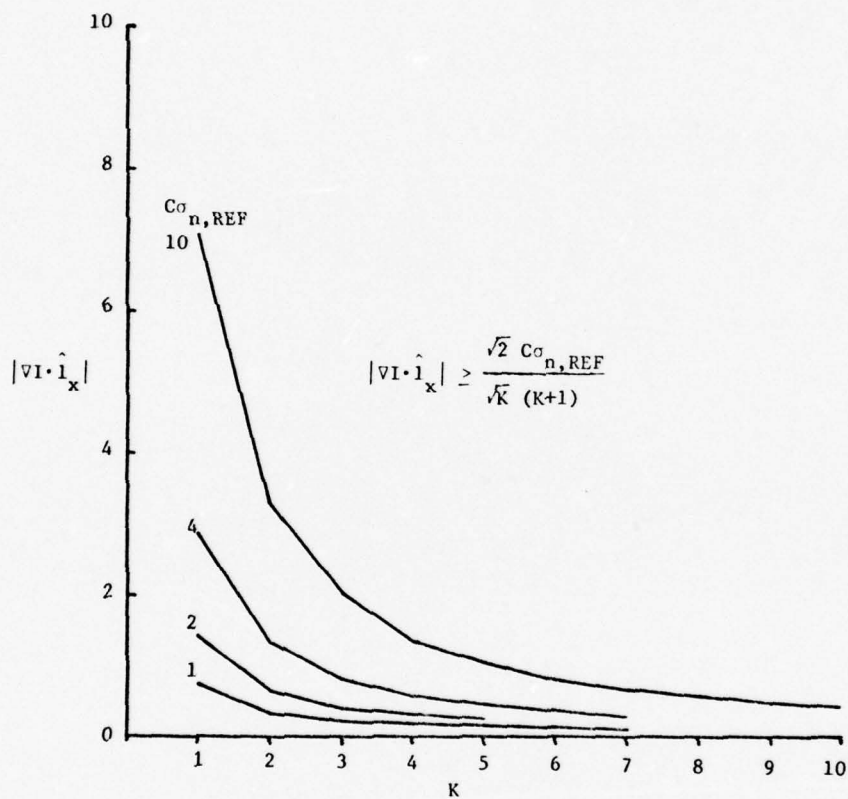We now have a gradient magnitude estimation algorithm with the

114

Figure 22.   Minimum Detectable Gradient Magnitude Components

115

desired characteristics. In the next section we will investigate the

performance of this estimator, the distribution of gradient magnitudes

in some actual images, and the improvement in $\gamma^2_{N,AVE}$ that can be

obtained by using the gradient magnitude histogram to select a

reference set.

4.3.2   Experimental Reference Selection Using Gradient

Magnitudes

Figure 23 through Figure 25 shows the distribution of detected

gradient magnitudes for the three different images with selected values

of C and estimated values of σ_n.

Figure 26 through Figure 31 display the gradient magnitudes as

detected.

In these images, large detected values of gradient magnitude show up as

bright pixels. From this it is clear that in all three images there

are a few pixels with very high gradient magnitudes and a large

majority with relatively small values. The result is that the value of

$m^2_{N,AVE}$ (dy,dx) decreases rapidly as N increases. Figure 32, Figure 33,

and Figure 34 locate the pixels incorporated into the reference set for

N = 128.

Figure 35 shows the auto-distance function for a 32 by 32 block

of adjacent pixels from the reference image shown in Figure 63.

The upper right corner of the reference set was located at row

43, column 43. When the gradient magnitude detection algorithm was

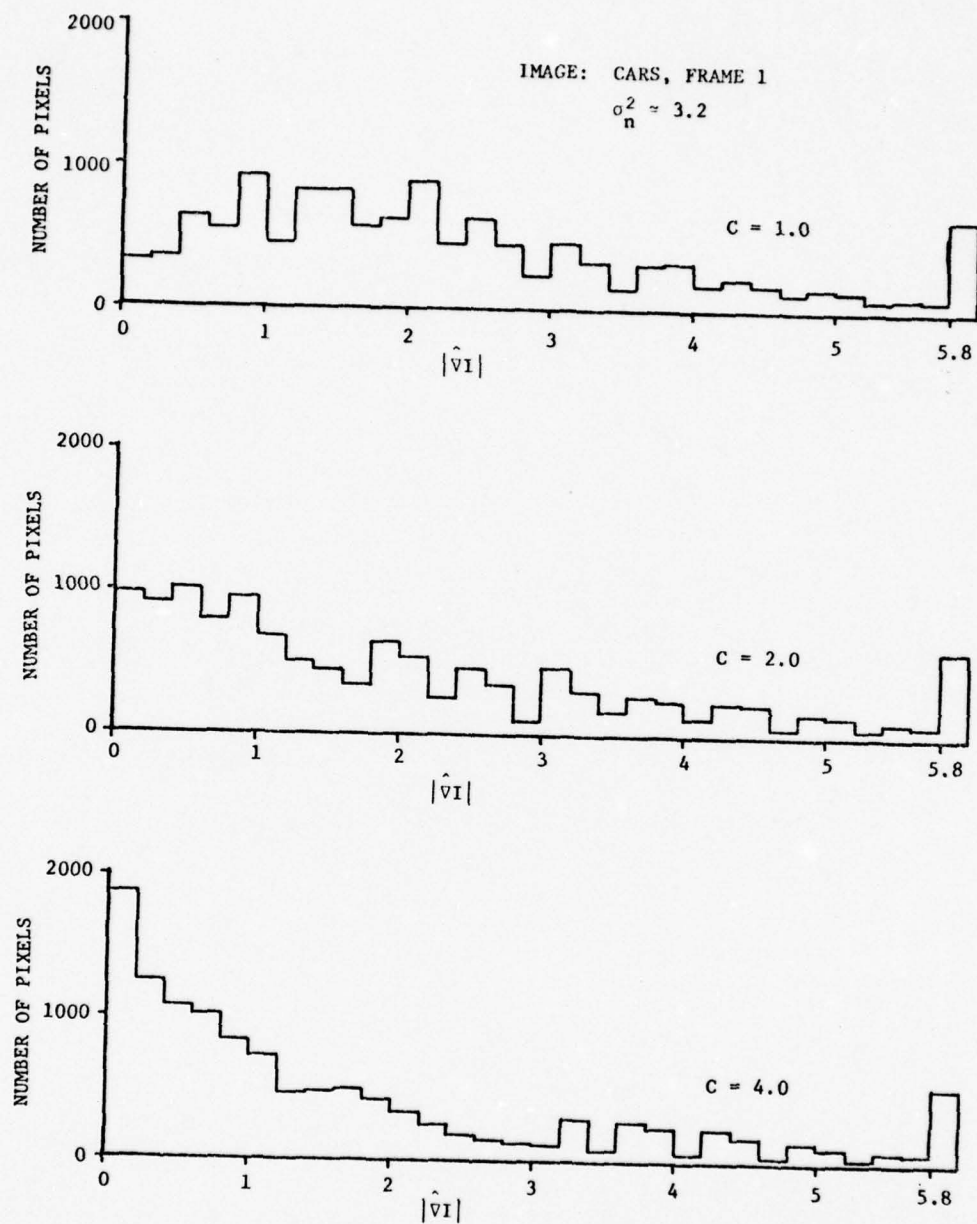used to select the reference set, a very significant increase in

116

IMAGE: CARS, FRAME 1

$\sigma_n^2 \simeq 3.2$

C = 1.0

NUMBER OF PIXELS

2000

1000

0

0   1   2   3   4   5   5.8

$|\hat{\nabla} I|$

NUMBER OF PIXELS

2000

1000

0

C = 2.0

0   1   2   3   4   5   5.8

$|\hat{\nabla} I|$

NUMBER OF PIXELS

2000

1000

0

C = 4.0

0   1   2   3   4   5   5.8
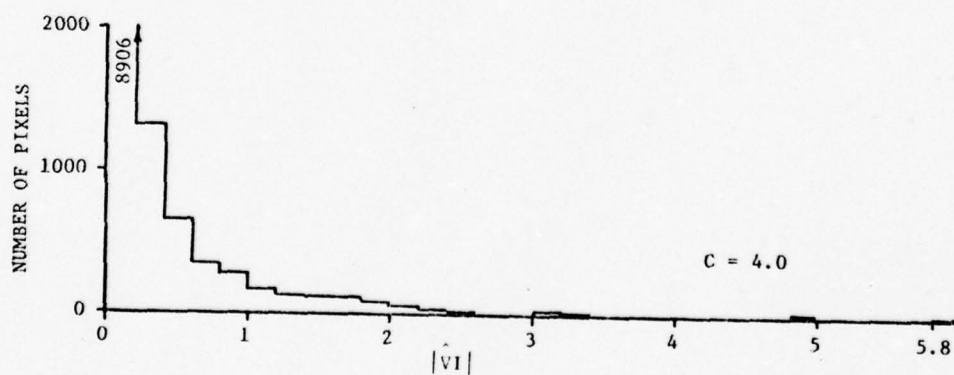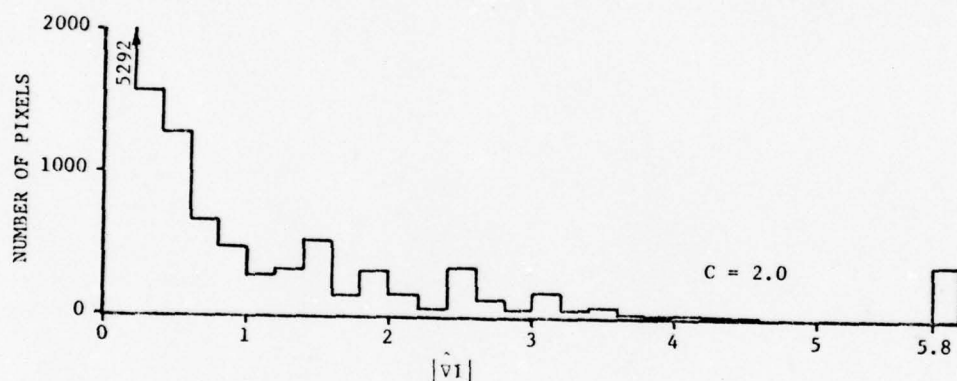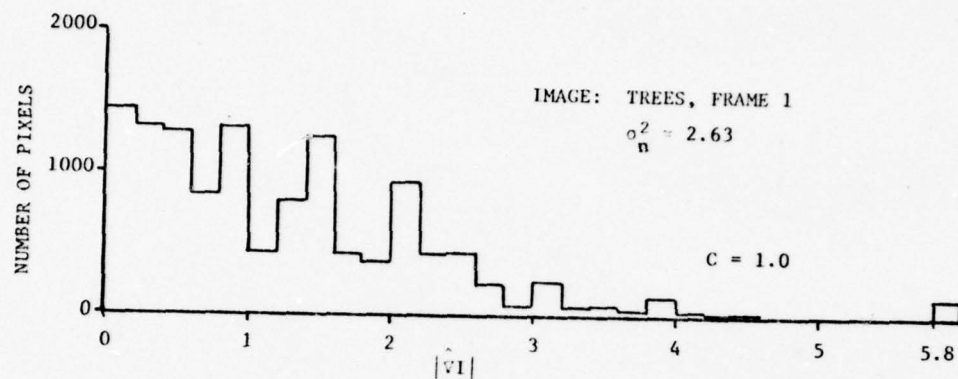
$|\hat{\nabla} I|$

Figure 23. Gradient Distribution for CARS
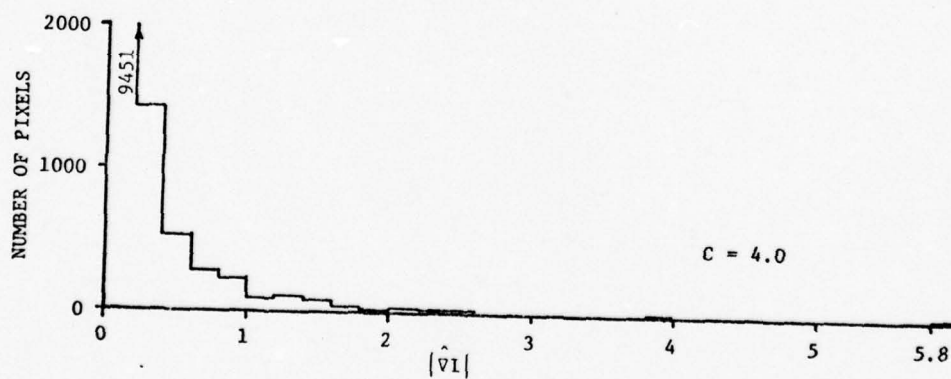
117

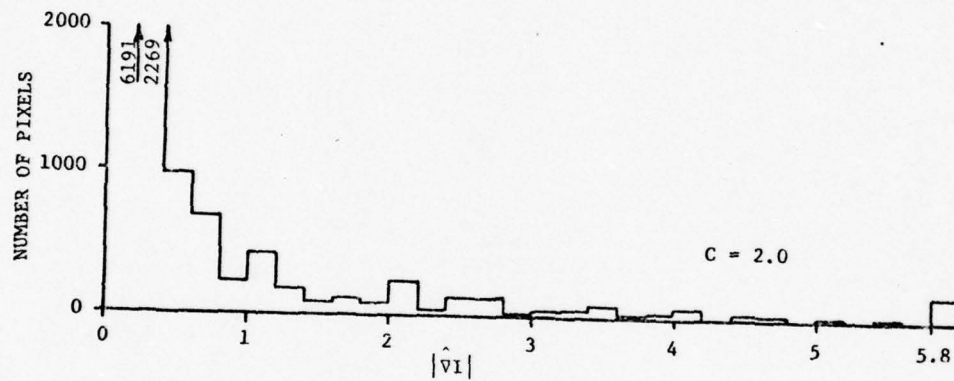Figure 24. Gradient Distribution for TREES
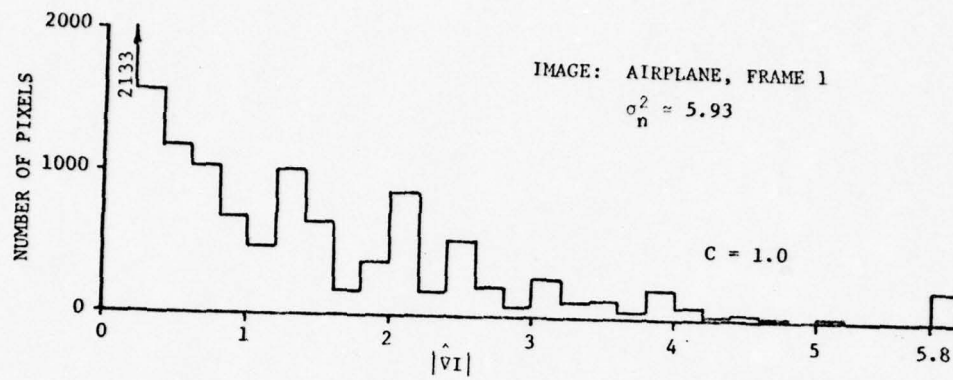
Figure 25. Gradient Distribution for AIRPLANE

119
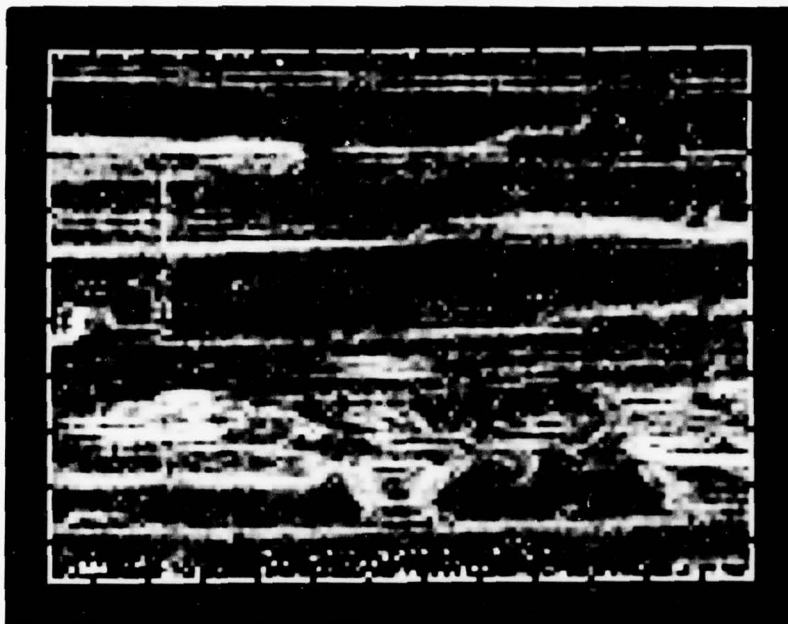
C = 1
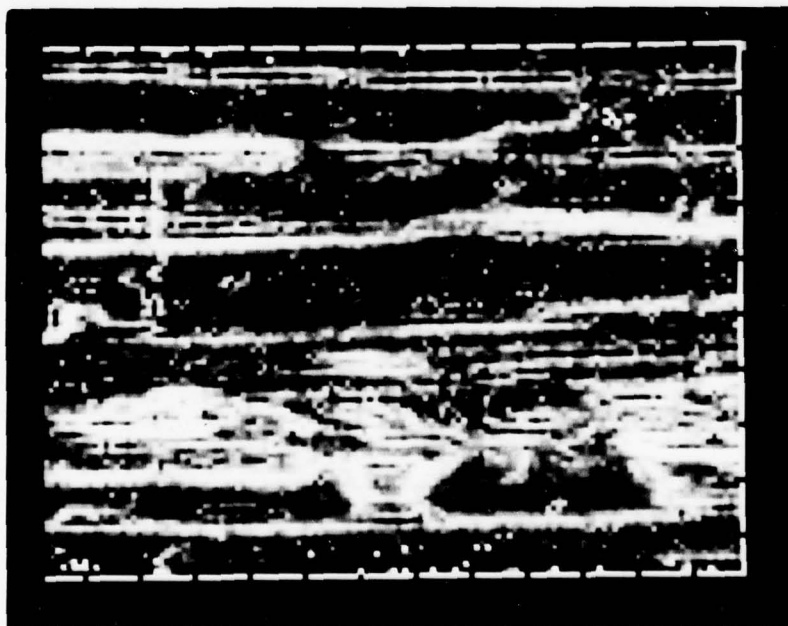
C = 2

Figure 26. Gradient magnitude image of CARS for C = 1 and C = 2

Figure 27. Gradient magnitude image of CARS for C = 4

121

C = 1

C = 2

Figure 28. Gradient magnitude image of TREES for C = 1 and C = 2

122

Figure 29. Gradient magnitude image of TREES for C = 4

C = 1

C = 2
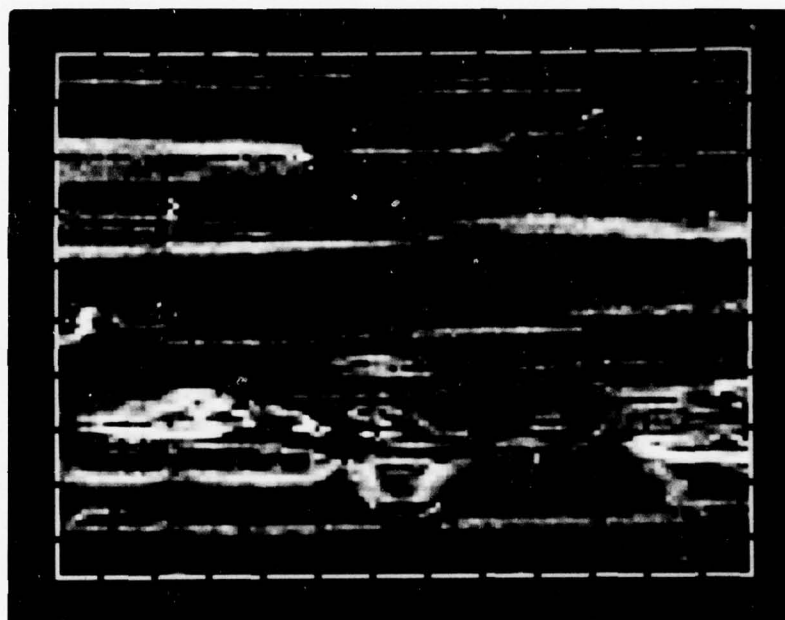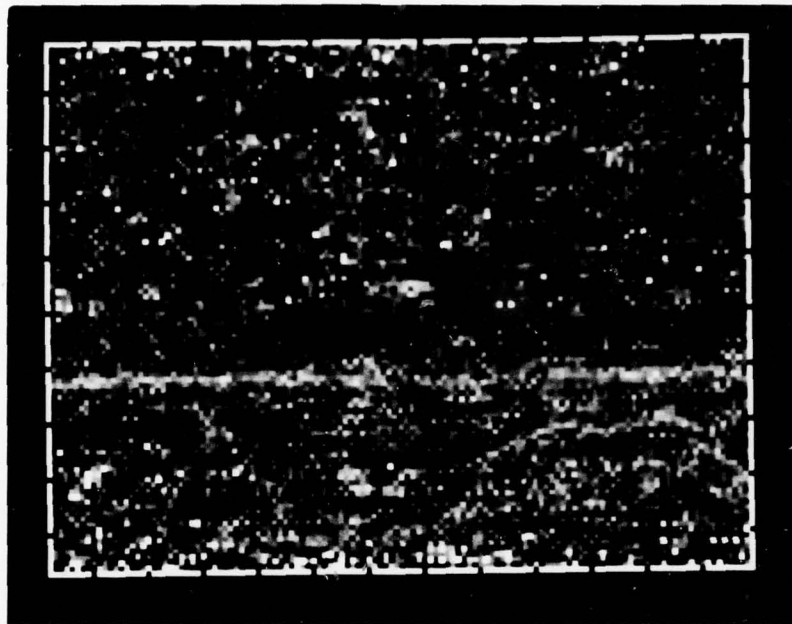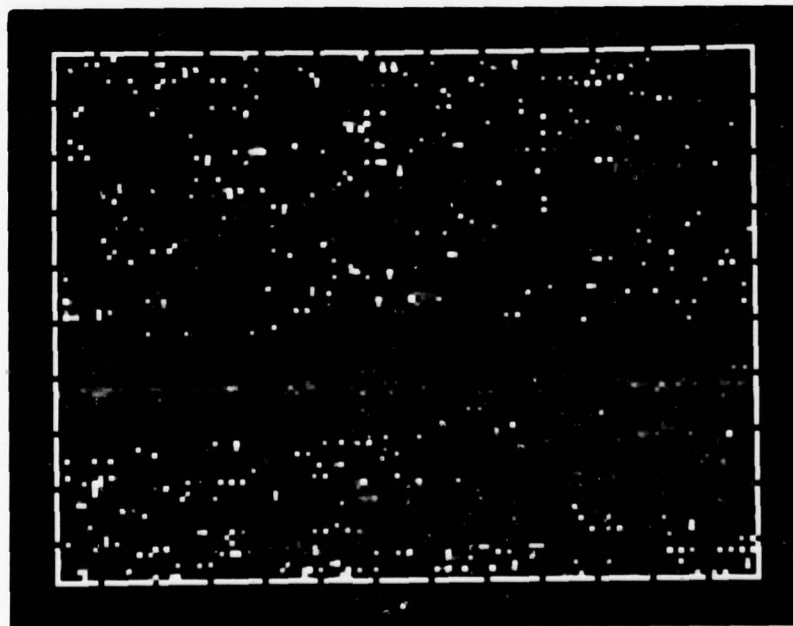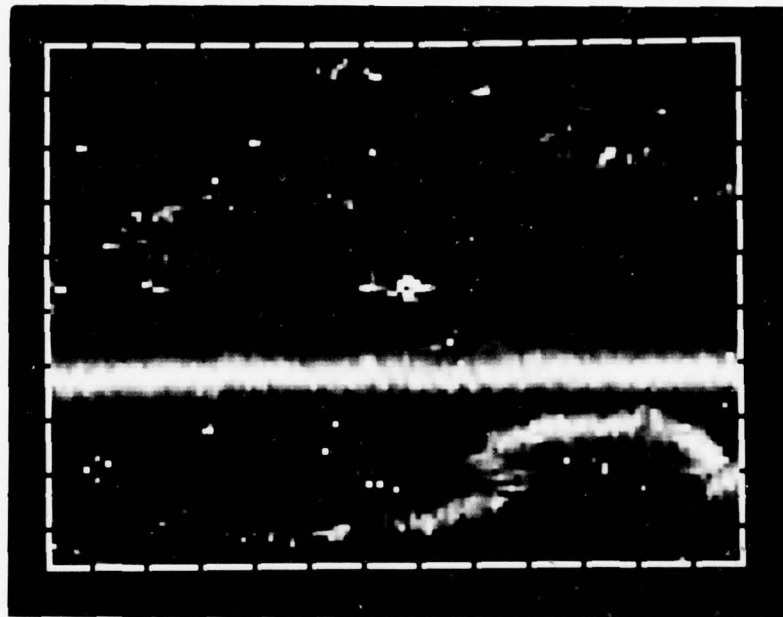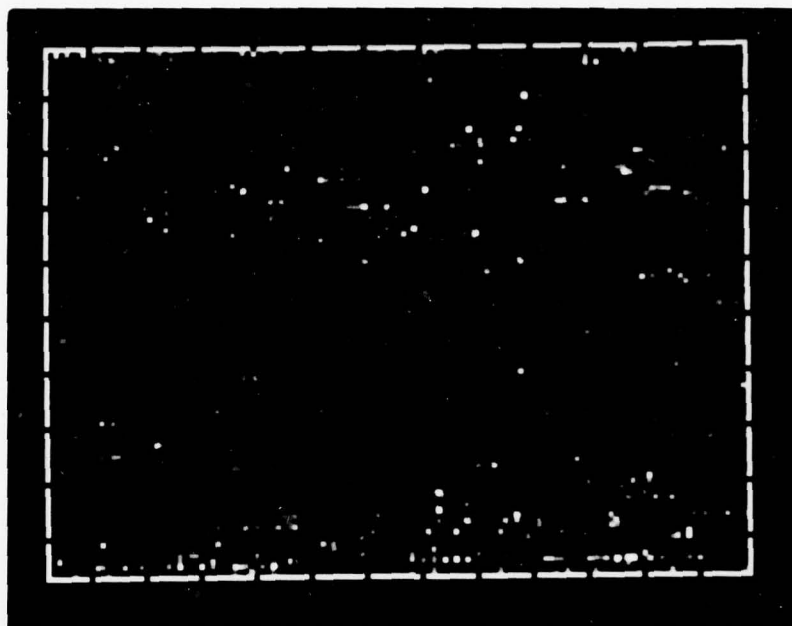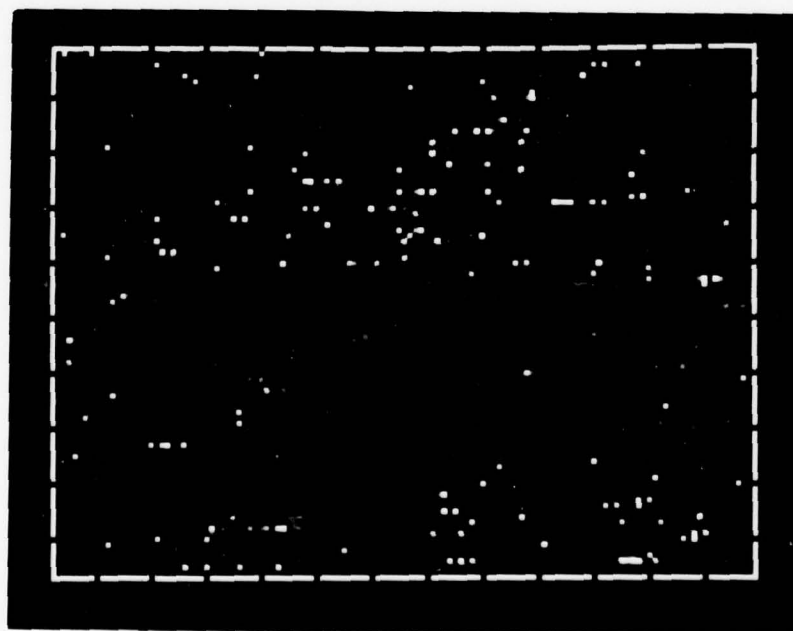
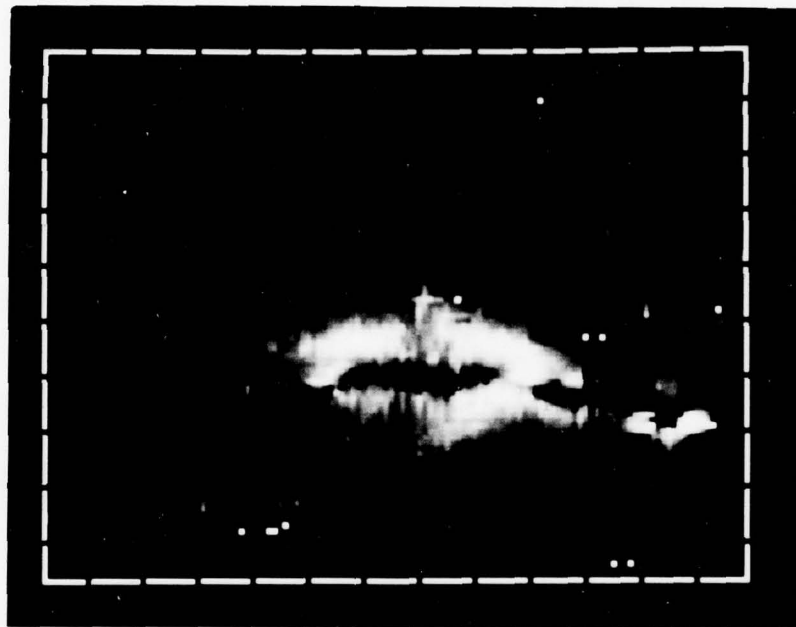Figure 30. Gradient magnitude image of AIRPLANE for C = 1 and C = 2

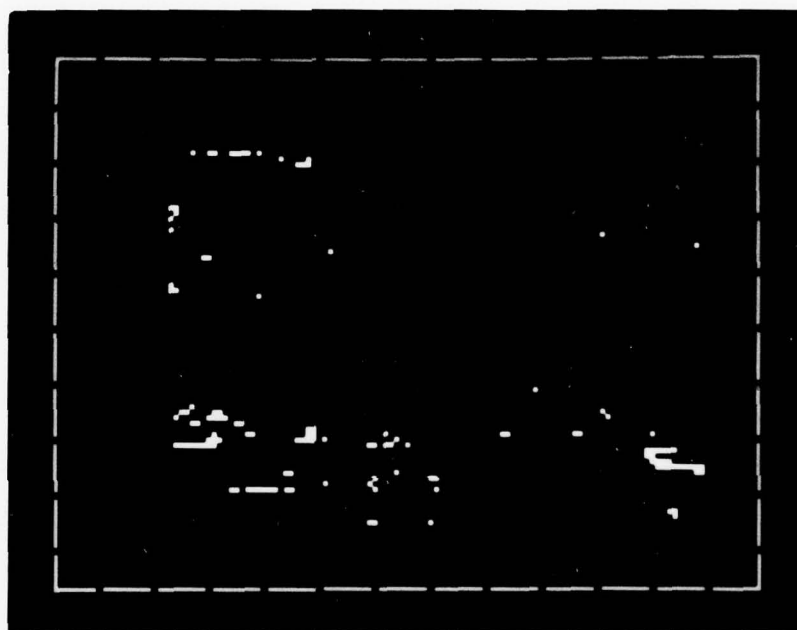Figure 31. Gradient magnitude image of AIRPLANE for C = 4

Figure 32. Reference set for CARS with N = 128
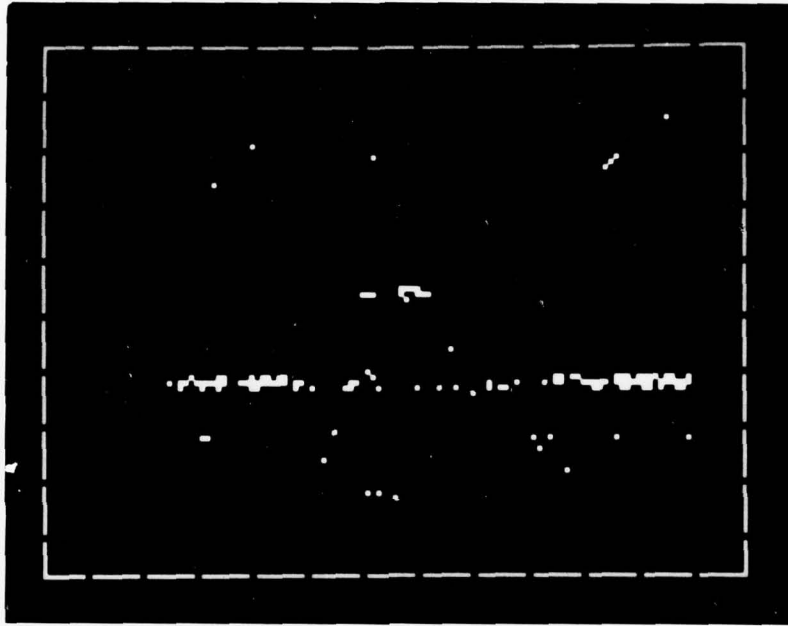
Figure 33. Reference set for TREES with N = 128
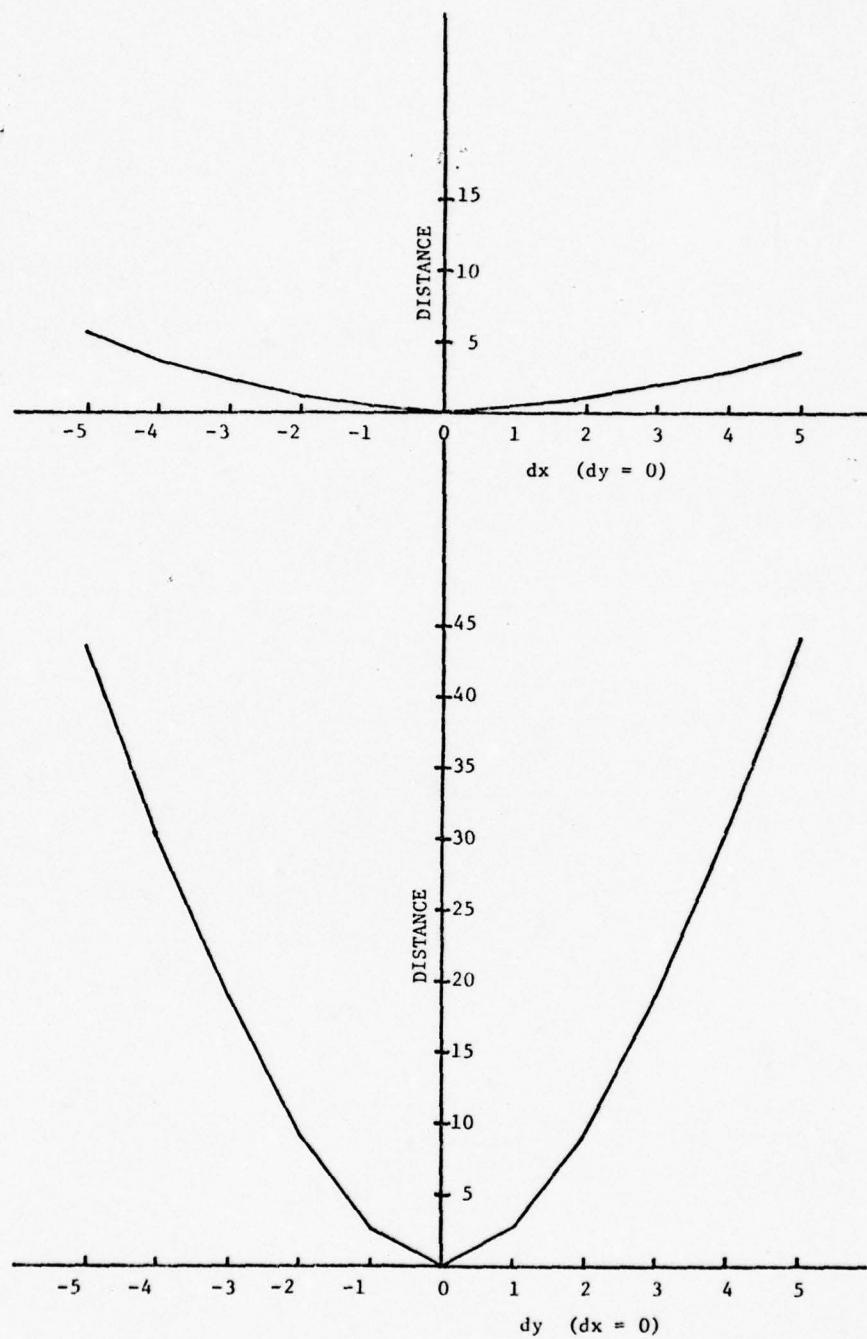
Figure 35. Auto-Distance Function for a 32 by 32
Block of Contiguous Pixels

average signal strength was realized. Figure 36 illustrates the improvement that was realized when the 1024 pixels with maximum gradient magnitudes were used to form the reference set. When N was reduced from 1024 to 128, a further increase in average signal strength was realized (see Figure 37).

In any tracking system which uses a single processing element to perform all of the computations associated with evaluating $d^2 (.,.)$, there will be a trade-off to be made between the number of trial registrations that can be attempted and the number of pixels to be carried in the comparison set. If the interframe time and the target dynamics relative to the sensor optical axis are known, a maximum search area size can be determined. Search area size, in combination with pixel spacing and processor speed, leads directly to an upper bound on values for N. At this point, specification of a maximum allowable probability of error will determine the minimum acceptable signal-to-noise ratio. If the combination of scene and sensor cannot provide the required signal-to-noise ratio, a faster processor or a smaller search area is indicated.

4.4   Summary

In Chapter 4 we have assumed that the reference image was given, and dealt with three techniques to increase the effective tracking signal-to-noise ratio: first, by using non-uniform weights for the norm; second, by reducing the random noise component of raw data imagery in regions of low contrast; and, third, by selecting the reference set in a way which increases its average signal strength. In the next chapter we will consider how to get a good reference image.

130
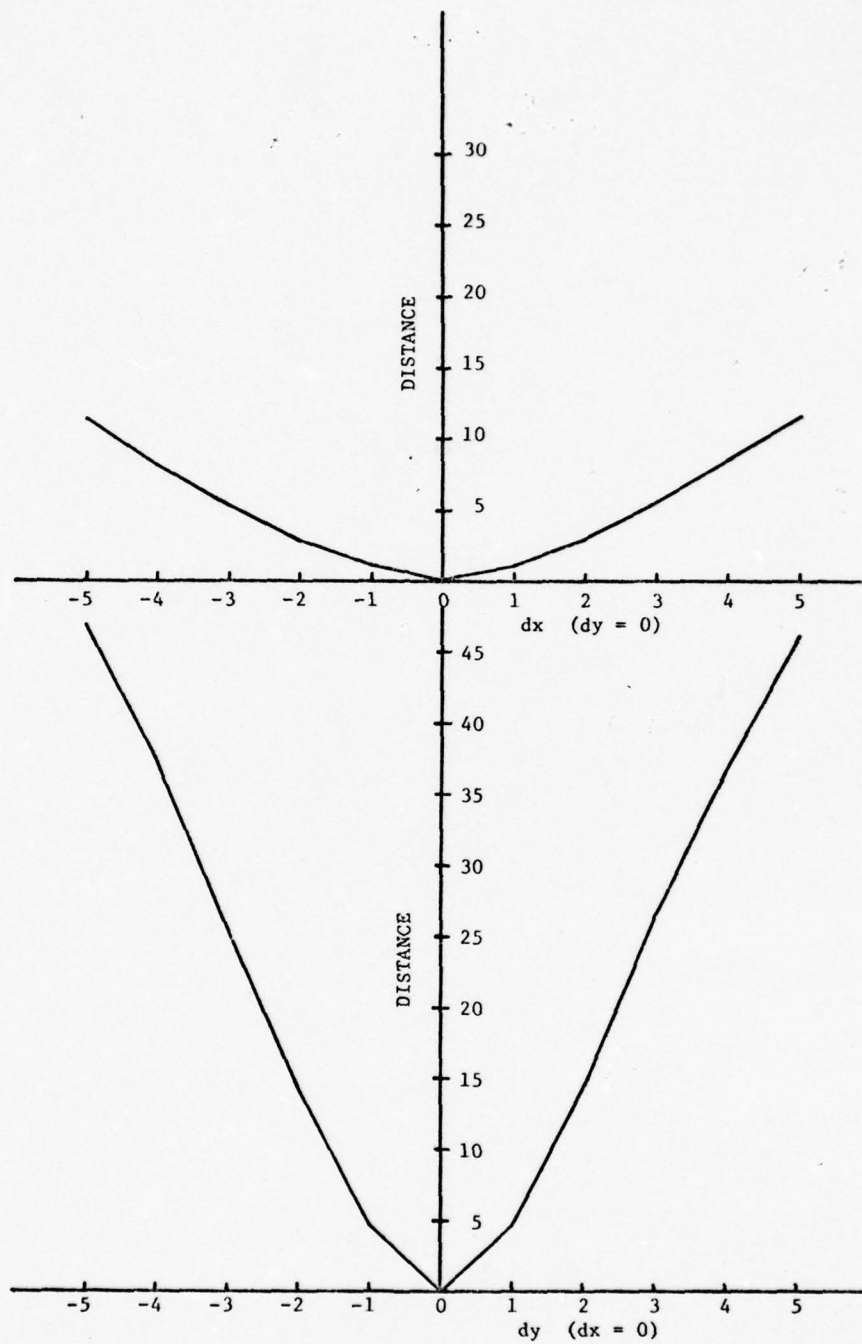
Figure 36.   Auto-Distance Function for a Reference Set
            Selected Using the Adaptive Gradient Detector (N=1024)

131
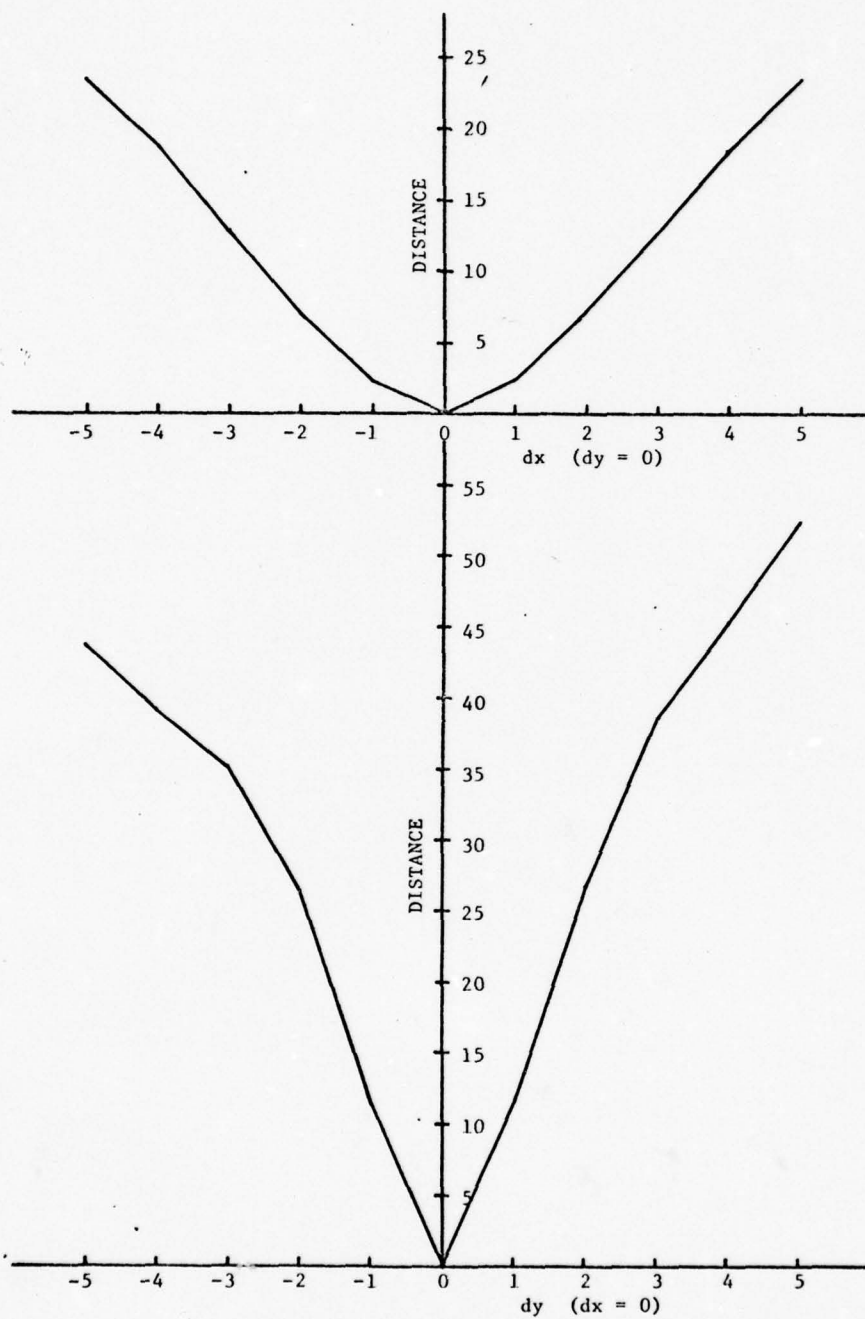
Figure 37. Auto-Distance Function for a Reference Set Selected
Using the Adaptive Gradient Detector (N=128)

132

Chapter 5

REFERENCE IMAGE ESTIMATION

In this chapter, we develop an adaptive Kalman filter to perform the reference-image-update task, prove that it is stable, and demonstrate the performance of the filter.

In Section 3.10 we investigated the sensitivity of the probability of error to both the signal and noise components of the image. While the $m_i(dy,dx)$ are a function of the sensor and the scene, the noise is a function of the sensor and the processing that is performed on the received images. If the signal-to-noise ratio at the input to the similarity detection process is maximized, then the performance of the system, as measured by the probability of error, is dependent on the reference-image-update process to minimize $\sigma_{n,REF}^2$.

There are also benefits to the reference-set selection process when $\sigma_{n,REF}^2$ is reduced since the gradient estimator performance is also dependent on this noise variable.

5.1 Adaptive Kalman Filter

The classic formulation of the Kalman filter assumes a complete a priori knowledge of the process and the measurement noise statistics. In most practical applications these statistics are inexactly known. The use of incorrect a priori statistics can result in a Kalman filter

133

which has large estimation errors or which may even be divergent. The

purpose of an adaptive filter is to reduce these errors by modifying

the filter to adapt it to the real data.

At this point we digress for a moment to review the general

Kalman filtering problem, we will then establish the the equivalence

between the conventional Kalman filter notation and its specific

application to the sequential image tracking problem and develop the

estimation procedure to be used to obtain the process and measurement

noise statistics. For a more detailed review of Kalman filter theory

see Gelb [12]. The particular approach to be followed in developing

the adaptive filter largely follows Mehra [25]. Let

$$x_{i+1} = \bigoplus x_i + u_i \qquad\qquad (5.1)$$

$$z_i = H x_i + v_i \qquad\qquad (5.2)$$

where x_i is the state vector, $\bigoplus$ is the state transition matrix, u_i is

the process noise vector which induces changes in x_i, z_i is the

measurement vector, H is the measurement matrix, and v_i is the

measurement noise vector. Both u_i and v_i are assumed to be zero-mean,

uncorrelated Gaussian sequences with

$$E[u_i] = 0 \qquad\qquad (5.3)$$

134

$$E[u_i \, u_j^T] = Q \, \delta_{ij} \qquad (5.4)$$

$$E[v_i] = 0 \qquad (5.5)$$

$$E[v_i \, v_j^T] = R \, \delta_{ij} \qquad (5.6)$$

where δ_{ij} is the Kronecker delta function and Q and R are bounded positive definite matrices. Let $\hat{x}_{i/j}$ be an estimate of x_i based on the observation set Z_j where

$$Z_j = \{z_1, z_2, \ldots, z_j\} \qquad (5.7)$$

Let $P_{i/j}$ be the covariance of the estimation error based on Z_j

$$P_{i/j} = E\left([x_i - \hat{x}_{i/j}] \, [x_i - \hat{x}_{i/j}]^T \right) \qquad (5.8)$$

When Q and R are known, the minimum variance linear estimator is given by the Kalman filter of the form

$$\hat{x}_{i+1/i} = \Phi \, \hat{x}_{i/i} \qquad (5.9)$$

135

$$\hat{x}_{i/i} = \hat{x}_{i/i-1} + K_i [z_i - H_i \hat{x}_{i/i-1}] \tag{5.10}$$

$$K_i = P_{i/i-1} H^T (H P_{i/i-1} H^T + R)^{-1} \tag{5.11}$$

$$P_{i/i} = (I - K_i H) P_{i/i-1} \tag{5.12}$$

$$P_{i+1/i} = \Phi P_{i/i} \Phi^T + Q \tag{5.13}$$

where K_i is the Kalman gain, and $\nu_i = z_i - Hx_i$ is called the innovation sequence. For the sequential image estimation problem both Q and R will be estimated from measurements made during the reference image update process.

The following equivalences establish the relationships between the notation of conventional Kalman filter theory and the particular variables of the image tracking problem as used in the preceding chapters. We let

$$x = Is, \text{ the underlying image} \tag{5.14}$$

$$\hat{x} = Ir, \text{ the reference image} \tag{5.15}$$

136

$$z = Id, \text{ the data image} \tag{5.16}$$

$$H = \bar{\Phi} = I, \text{ the identity matrix} \tag{5.17}$$

$$R = \sigma^2_{n,DATA} I, \text{ the measurement noise covariance matrix}$$

$$\tag{5.18}$$

$$Q = q^2 I, \text{ the process noise covariance matrix}$$

$$\tag{5.19}$$

$$P_{i/i-1} = \sigma^2_{n,REF} I, \text{ the estimation error covariance matrix}$$

$$\tag{5.20}$$

where all of the identity matrices are of dimension $M_R M_C$ (the number of pixels in the reference image). We employ a covariance-matching technique to determine appropriate values for $\sigma^2_{n,DATA}$ and q^2. Note that while the underlying image is strictly positive (or zero), the state model allows for negative state vector components. To the extent that this does not represent the true situation with real images, the filter may produce suboptimal results. The expected value of the innovation sequence is [25]

137

$$E[\nu_i \nu_i^T] = E[(z_i - H x_{i/i-1}) (z_i - H x_{i/i-1})^T]$$

$$= P_{i/i-1} + R$$

$$= (\sigma_{n,REF}^2 + \sigma_{n,DATA}^2) I \qquad (5.21)$$

Any detected deviation above this value is taken as an indication that the filter is not optimal (in the sense of minimum variance) and that q^2 and $\sigma_{n,DATA}^2$ should be adjusted to bring the filter back toward optimal performance. The reference image update filter will maintain estimates of both $\sigma_{n,DATA}^2$ and $\sigma_{n,REF}^2$ and use the difference image associated with the minimum distance registration as the innovation sequence.

Since the difference image contains a large number of pixels, the sample statistics for the difference image should closely approximate the true underlying statistics, i.e. the bias and variance of the sample statistics will be small. The sample statistic that will be used in the estimation of Q and R is the difference image sample variance v_i.

138

$$v_i = \frac{1}{M_D} \sum_k \sum_j D_i^2(dy,dx,k,j) - \left\{ \frac{\sum_k \sum_j D_i^2(\widehat{dy},\widehat{dx},k,j)}{M_D} \right\}^2$$

(5.22)

where M_D is the number of pixels in the difference image associated with the minimum distance.

$$M_D = \left[MIN\{M_R, M_R-dy\} - MAX\{1, dy\} \right] \times$$

$$\left[MIN\{M_C, M_C-dx\} - MAX\{1, dx\} \right]$$

(5.23)

The mean and variance of v_i are

$$E[v_i] = \sigma_{n,REF}^2(i-1) + \sigma_{n,DATA}^2(i-1)$$

(5.24)

and

$$var[v_i] = \frac{2[\sigma_{n,REF}^2(i-1) + \sigma_{n,DATA}^2(i-1)]}{M_D - 1}$$

(5.25)

While v_i is a biased estimate (N_D should be smaller by one to be

unbiased), 5.21 is prefered over the unbiased estimate since it gives

less mean-square error [15]. We will use $v_i - \sigma^2_{n,REF}(i-1)$ as a measure

of change in $\hat{\sigma}^2_{n,DATA}$, and define a time constant β $(0 < \beta < 1)$ for

changes in $\hat{\sigma}^2_{n,DATA}$.

$$\hat{\sigma}^2_{n,DATA}(i) = \beta \hat{\sigma}^2_{n,DATA}(i-1) + (1-\beta)[v_i - \hat{\sigma}^2_{n,REF}(i-1)]$$

$$(5.26)$$

The basis for this restriction on the rate of change of $\sigma^2_{n,DATA}$ is the

assumption that sensor noise variance is a function of parameters which

change relatively slowly compared to the sensor frame rate. In a

vidicon sensor, it could be the faceplate temperature or target

voltage. In other sensor types, other noise sources respond to the

environment with finite time response. The remainder of the difference

between v_i and the filter estimates of $\sigma^2_{n,DATA}$ and $\sigma^2_{n,REF}$, denoted

T_2, is attributed to change in the underlying image Is and is assigned

to q^2_i to increase the filter estimate of $\sigma^2_{n,REF}$ prior to the

calculation of the next Kalman gain.

$$T_2 = v_i - \hat{\sigma}^2_{n,REF}(i-1) - \hat{\sigma}^2_{n,DATA}(i-1) \qquad (5.27)$$

Since Q and R must be positive definite matrices, a precautionary restriction is placed on q_i^2 and on $\hat{\sigma}^2_{n,DATA}$ (see (5.33))

$$q_i^2 = MAX\{\, 0,\ T_2 \,\} \qquad (5.28)$$

To initialize the filter, we take the first data image as the first reference image since there is no better information available about Is, and for $\hat{\sigma}^2_{n,REF}$, we use $\hat{\sigma}^2_{n,DATA}$, an a priori guess at the variance of the noise component of Id.

The full set of adaptive filter equations is summarized as follows in terms of the variables unique to this problem:

Initialization

$$\hat{\sigma}^2_{n,REF}(1) = \sigma^2_{n,DATA}(1) \qquad (5.29)$$

141

Update

$$\bar{K}_i = \frac{\hat{\sigma}^2_{n,REF}(i-1)}{\hat{\sigma}^2_{n,REF}(i-1) + \hat{\sigma}^2_{n,DATA}(i-1)} \qquad (5.30)$$

$$Ir_i(k,j) = Ir_{i-1}(k-dy,\ j-dx) + \bar{K}_i\ [Id_{i-1}(k,j) \qquad (5.31)$$
$$- Ir_{i-1}(k-dy,\ j-dx)]$$

Propagation

$$v_i = \frac{1}{M_D}\sum_k \sum_j D_i^2(dy,dx,k,j) - \left[\frac{\sum_k \sum_j D_i^2(dy,dx,k,j)}{M_D}\right]^2 \qquad (5.32)$$

$$T_1 = \hat{\sigma}^2_{n,DATA}(i-1) + (1-\beta)[v_i - \hat{\sigma}^2_{n,REF}(i-1)] \qquad (5.33)$$

$$\hat{\sigma}^2_{n,DATA}(i) = MAX\{\ 0,\ T_1\ \} \qquad (5.34)$$

$$T_2 = v_i - \hat{\sigma}^2_{n,REF}(i-1) - \hat{\sigma}^2_{n,DATA}(i) \qquad (5.35)$$

142

$$q_i^2 = \text{MAX}\{\, 0,\ T_2\, \} \tag{5.36}$$

$$\hat{\sigma}_{n,REF}^2(i) = \frac{\hat{\sigma}_{n,REF}^2(i-1)\, \hat{\sigma}_{n,DATA}^2(i)}{\hat{\sigma}_{n,REF}^2(i-1) + \hat{\sigma}_{n,DATA}^2(i)} + q_i^2 \tag{5.37}$$

In the next section we will analyze the stability of this algorithm.

5.2   Filter Stability Analysis

. In the design of an adaptive Kalman filter, because of the adhoc nature of the covariance matching process, the question of stability must be addressed. In this section, we show that the filter is stable except during periods when the observed difference image variance indicates that the filter estimate of $\sigma_{n,DATA}^2 + \sigma_{n,REF}^2$ is too low. During these periods the filter enters an unstable region of operation, increasing $\hat{\sigma}_{n,DATA}^2 + \hat{\sigma}_{n,REF}^2$ until the sum is once again in agreement with the observed data.

Eliminating q_i^2 from (5.26) and (5.36) and letting

143

$$K_i^+ = \frac{\hat{\sigma}_{n,REF}^2 (i-1)}{\hat{\sigma}_{n,REF}^2 (i-1) + \hat{\sigma}_{n,DATA}^2 (i)} \qquad (5.38)$$

we have recursive equations for $\hat{\sigma}_{n,DATA}^2$ and $\hat{\sigma}_{n,REF}^2$ with v as the only forcing function

$$\hat{\sigma}_{n,DATA}^2 (i) = \hat{\sigma}_{n,DATA}^2 (i-1) + (1-\beta)v_i - (1-\beta) \hat{\sigma}_{n,REF}^2 (i-1)$$

$$(5.39)$$

$$\hat{\sigma}_{n,REF}^2 (i) = [1-K_i^+] \hat{\sigma}_{n,REF}^2 (i-1) + v_i \qquad (5.40)$$

$$- \hat{\sigma}_{n,REF}^2 (i-1) - \hat{\sigma}_{n,DATA}^2 (i)$$

The constraints placed on the propagation equations define four potential regions of operation for the filter:

Region I

$$T_1 < 0$$

$$T_2 < 0$$

144

Region II

$$T_1 > 0$$

$$T_2 < 0$$

Region III

$$T_1 > 0$$

$$T_2 > 0$$

Region IV

$$T_1 < 0$$

$$T_2 > 0$$

From (5.32) and (5.34) we see that

$$T_1 \geqslant 0 \Rightarrow v_i > \hat{\sigma}^2_{n,REF}(i-1) - \frac{\beta}{1-\beta}\hat{\sigma}^2_{n,DATA}(i-1) \qquad (5.41)$$

and

$$T_2 > 0 \Rightarrow v_i > \hat{\sigma}^2_{n,REF}(i-1) + \hat{\sigma}^2_{n,DATA}(i-1) \qquad (5.42)$$

hence it is clear that $T_2 > 0$ is the more restrictive constraint, i.e.

145

$$T_2 > 0 \Rightarrow T_1 > 0$$

which precludes the possibility of operating in Region IV.
Incorporating the constraints associated with each region of operation
and writing the resulting equations in matrix form:

Region I

$$T_1 \leqslant 0 \qquad \hat{\sigma}^2_{n,DATA}(i) = 0 \qquad (5.43)$$

$$T_2 \leqslant 0 \implies q_i^2 = 0 \qquad (5.44)$$

therefore

$$\hat{\sigma}^2_{n,REF}(i) = 0 \qquad (5.45)$$

and

$$s_i = 0 \qquad (5.46)$$

where

$$s_i = \begin{bmatrix} \hat{\sigma}^2_{n,REF}(i) \\[2ex] \hat{\sigma}^2_{n,DATA}(i) \end{bmatrix} \qquad (5.47)$$

147

Region II

$$T_1 > 0 \Rightarrow \hat{\sigma}^2_{n,DATA}(i) = \beta \, \hat{\sigma}^2_{n,DATA}(i-1)$$

$$+ (1-\beta)[v_i - \hat{\sigma}^2_{n,REF}(i-1)] \tag{5.48}$$

$$T_2 \leqslant 0 \Rightarrow q_i^2 = 0 \tag{5.49}$$

therefore

$$\hat{\sigma}^2_{n,REF}(i) = K_i^+ \, \hat{\sigma}^2_{n,DATA}(i)$$

$$= K_i^+ \beta \, \hat{\sigma}^2_{n,DATA}(i-1) + K_i^+ (1-\beta) v_i$$

$$- K_i^+ (1-\beta) \, \hat{\sigma}^2_{n,REF}(i-1) \tag{5.50}$$

and

$$s_i = \begin{bmatrix} -K_i^+(1-\beta) & K_i^+\beta \\ \\ -(1-\beta) & \beta \end{bmatrix} s_{i-1} + \begin{bmatrix} K_i^+(1-\beta) \\ \\ 1-\beta \end{bmatrix} v_i \tag{5.51}$$

Region III

$$T_1 > 0 \Rightarrow \hat{\sigma}^2_{n,DATA}(i) = \hat{\sigma}^2_{n,DATA}(i-1) \qquad (5.52)$$

$$+ (1-\beta)[v_i - \hat{\sigma}^2_{n,REF}(i-1)]$$

$$T_2 > 0 \Rightarrow q_i = v_i - \hat{\sigma}^2_{n,REF}(i-1) - \hat{\sigma}^2_{n,DATA}(i-1) \qquad (5.53)$$

therefore

$$\hat{\sigma}^2_{n,REF}(i) = K^+_i \hat{\sigma}^2_{n,DATA}(i) + v_i$$

$$- \hat{\sigma}^2_{n,REF}(i-1) - \hat{\sigma}^2_{n,DATA}(i-1)$$

$$= K^+_i \beta \hat{\sigma}^2_{n,DATA}(i-1) + K^+_i (1-\beta)v_i - K^+_i (1-\beta) \hat{\sigma}^2_{n,REF}(i-1)$$

$$+ v_i - \sigma^2_{n,REF}(i-1) - \hat{\sigma}^2_{n,DATA}(i-1)$$

149

$$= -[K_i^+(1-\beta)+1]\,\hat{\sigma}^2_{n,REF} + [K_i^+\beta-1]\,\hat{\sigma}^2_{n,DATA}(i-1)$$

$$+\,[K_i^+(1-\beta)+1]v_i \qquad (5.54)$$

and

$$s_i = \begin{bmatrix} -K_i^+(1-\beta)-1 & K_i^+\beta-1 \\[2mm] -(1-\beta) & \beta \end{bmatrix} s_{i-1} + \begin{bmatrix} K_i^+(1-\beta)+1 \\[2mm] 1-\beta \end{bmatrix} v_i \qquad (5.55)$$

In order for the filter to be stable, the homogeneous solution to the propagation equations must decay to zero. This requires that the eigenvalues of the propagation matrix lie inside the unit circle. In Region I, $s_i = 0$ is a degenerate case. K_{i+1} will be indeterminate and can be taken as either one, based on the prediction that the next data image will be perfect, or zero, based on the observation that the reference image is already perfect. In practice, this case will rarely occur, and when it does, both $\sigma^2_{n,DATA}$ and $\sigma^2_{n,REF}$ will be restored to nonzero values as soon as a nonzero v_i is observed; thus, the filter is stable in this region of operation.

For Regions II and III, we solve the characteristic equation for the eigenvalues of the propagation matrix and investigate the range of possible eigenvalues in each case.

150

Region II

$$\begin{vmatrix} -K_1^+(1-\beta) - & K_1^+\beta \\[2mm] -(1-\beta) & \beta-\lambda \end{vmatrix} = 0$$

$$= \lambda^2 - \beta\lambda + K_1^+(1-\beta)\lambda \qquad (5.56)$$

the solutions are $\lambda_1 = 0$ and $\lambda_2 = \beta - K_1^+(1-\beta)$. For λ_2 to satisfy $|\lambda_2| < 1$, there are two cases:

case 1

$$\beta - K_1^+(1-\beta) < 1 \qquad (5.57)$$

$$-K_1^+ < 1 \qquad (5.58)$$

which is always true, and

case 2

$$-\beta + K_1^+(1-\beta) < 1 \qquad (5.59)$$

151

$$K_i^+ (1-\beta) < 1 + \beta \qquad (5.60)$$

but

$$K_i^+ \leqslant 1 \text{ and } 1 - \beta < 1 \qquad (5.61)$$

so that

$$K_i^+ (1-\beta) < 1 < 1 + \beta \qquad (5.62)$$

and Region II operation is stable.

152

Region III

$$\begin{vmatrix} -K_i^+(1-\beta) - 1 - \lambda & K_i^+\beta - 1 \\ -(1-\beta) & \beta - 1 \end{vmatrix} = 0$$

$$= \lambda^2 + (1-\beta)(1 + K_i^+)\lambda - 1 \qquad (5.63)$$

the solutions are

$$\lambda_1 = -\frac{(1-\beta)(1 + K_i^+)}{2} + \sqrt{1 + \left[\frac{(1-\beta)(1 + K_i^+)}{2}\right]^2} \qquad (5.64)$$

and

$$\lambda_2 = -\frac{(1-\beta)(1 + K_i^+)}{2} - \sqrt{1 + \left[\frac{(1-\beta)(1 + K_i^+)}{2}\right]^2} < -1 \qquad (5.65)$$

Region III is an unstable region of operation for the filter, but the only time that the filter will operate in this region is when there is evidence (from the innovation sequence) that the sum of $\hat{\sigma}^2_{n,DATA}$ and

153

$\hat{\sigma}^2_{n,REF}$ is too small and should be increased to match the observed

sample variance of the difference image.

This might occur if the frame-to-frame translation error

exceeded the radius of the search region. In this case the correct

registration would not be one of the trial registrations and the

minimum distance would be greater than the predicted value. This event

is indicative of a change in the underlying image with respect to the

reference image and could be interpreted as a manifestation of a loss

of track. Thus, the filter will operate in Region III until T_2 becomes

negative; at which time, the filter reverts to stable operation in

Region II.

In Chapter 6 a simulation result will illustrate this

characteristic rapid adaptation of the filter to a loss of lock

condition and the ability of an integrated tracking algorithm to

reacquire the target automatically.

5.3   Kalman Filter Performance

The performance of the reference-image update process is

strongly dependent on the ability of the similarity detection algorithm

to correctly register the incoming data image. If the location of the

best match does not correspond to the correct registration, the filter

will incorporate the resulting error into the reference image and

increase the reference-image noise variance.

Figure 38 illustrates the ability of the adaptive Kalman filter

to correctly estimate the data-image noise variance, even though the
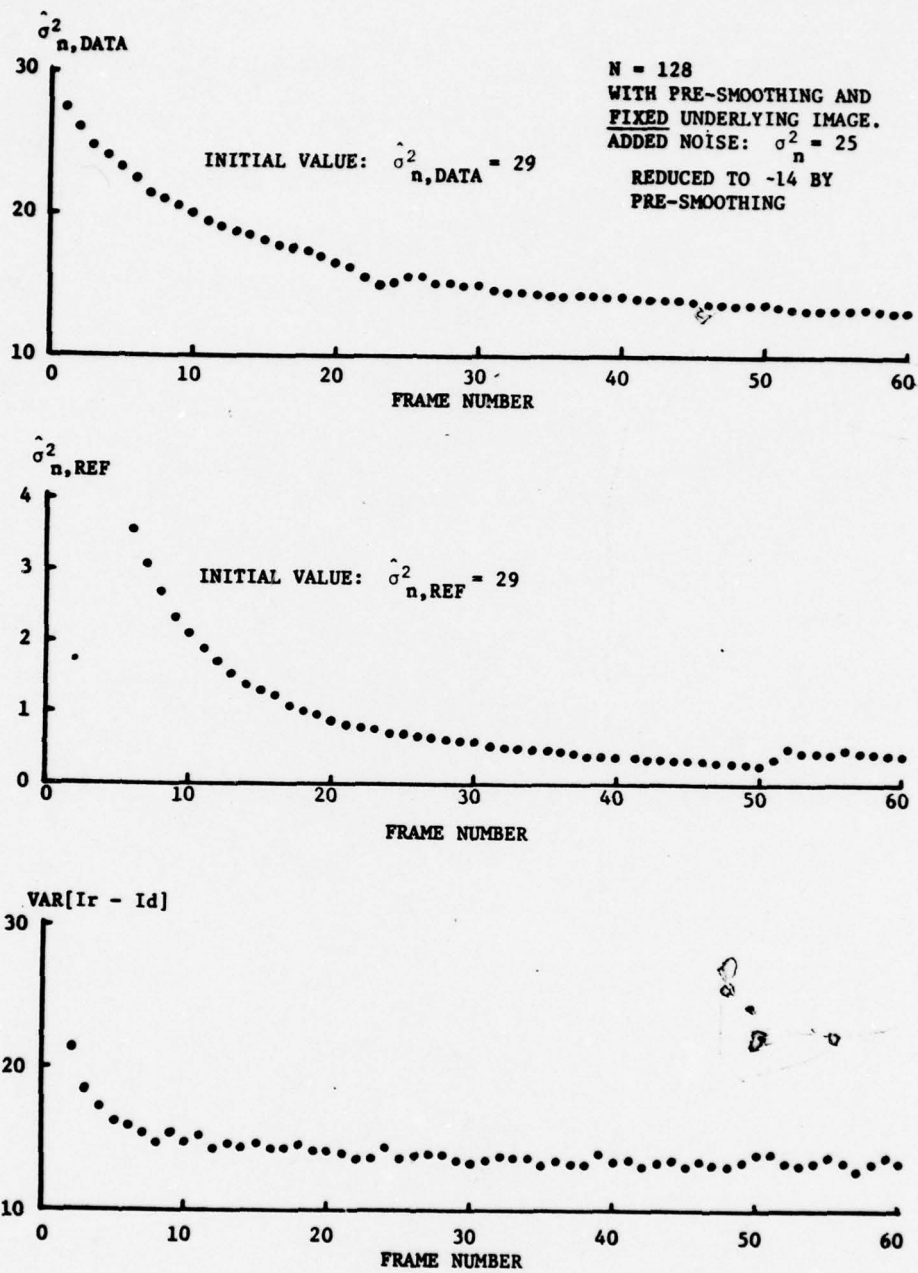
initial estimate is considerably in error.

154

Figure 38.   Kalman Filter Variance Estimates and
Residual Error Variance

155

In this case, the data image sequence was generated by
corrupting the image of Figure 63 with an uncorrelated Gaussian noise
sequence with $\sigma^2_{n,DATA}$ = 25. The initial value for $\sigma^2_{n,DATA}$ was 29 and
β was set to have a time constant of 15 frames. Even though the
incoming data has a noise component with variance 25, application of
the non-linear peak elimination filter as a presmoother results in a
filter estimate of the reference-image noise variance of only 14. This
agrees quite closely to the approximately 50% reduction in noise
variance which was realized when the non-linear peak elimination filter
was applied to a pure noise image. Operating at this noise level with
N = 128, the similarity detection process does make errors, as shown by
Figure 39.

After 60 frames, the tracker has built up an error of 2 pixels
in the horizontal direction, and 1 pixel in the vertical direction.
Note that most of the errors accumulate during the initialization
transient and before the Kalman filter has had time to reduce the
reference image noise variance.

In Section 4.2 the nonlinear peak elimination prefilter was
developed to reduce the random noise in regions of the image with low
gradient magnitudes. Figure 40 shows the difference image distribution
with and without the prefilter in use.

From this, it seems clear at while the prefilter does reduce
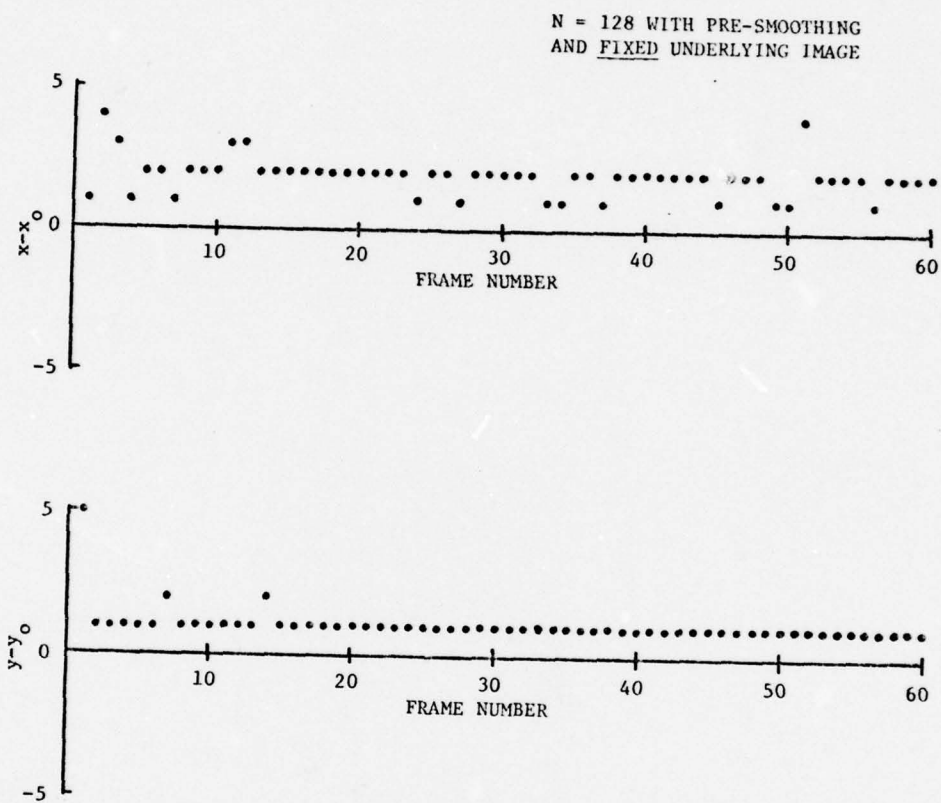the difference-image variance, the characteristic shape of the
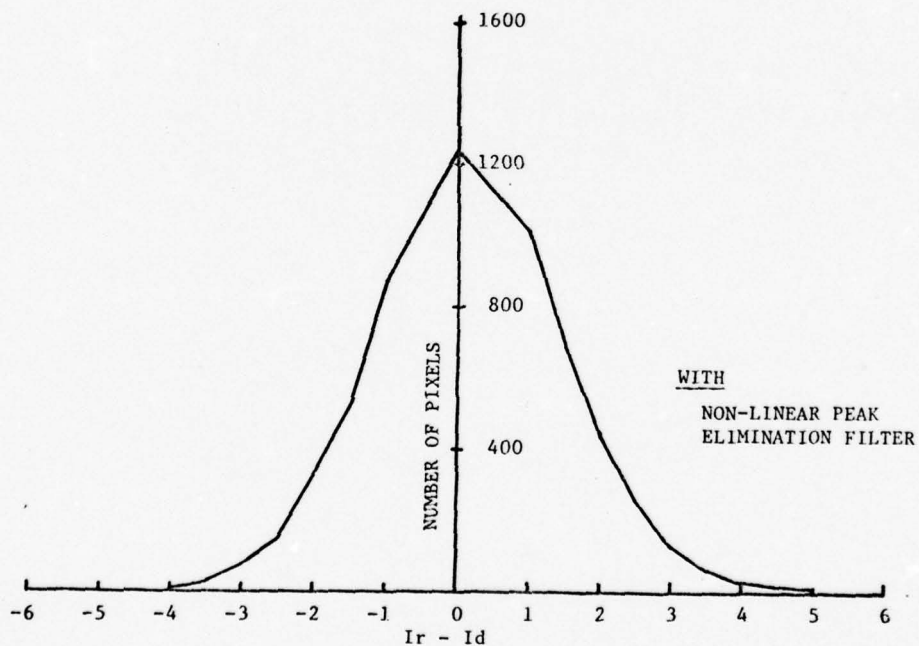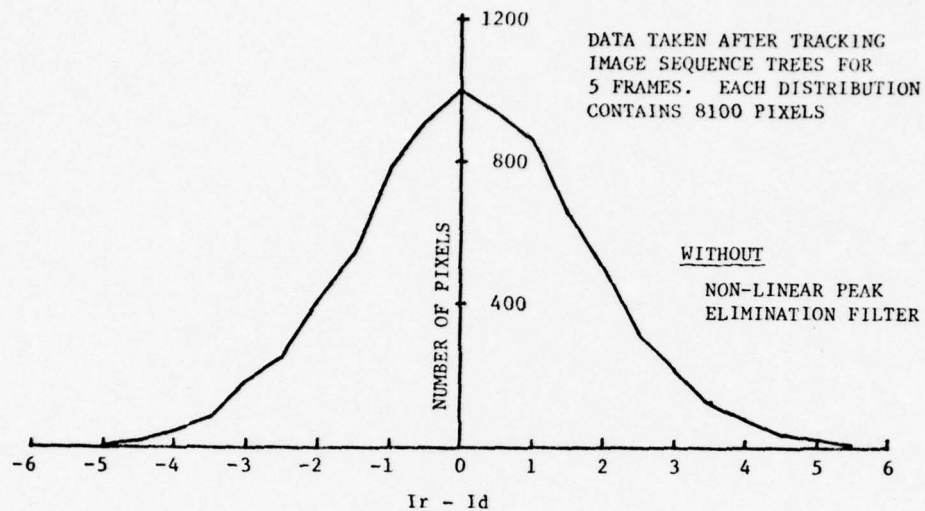distribution is preserved.

156

Figure 39. Accumulated Tracking Error

157

DATA TAKEN AFTER TRACKING
IMAGE SEQUENCE TREES FOR
5 FRAMES.   EACH DISTRIBUTION
CONTAINS 8100 PIXELS

WITHOUT

NON-LINEAR PEAK
ELIMINATION FILTER

WITH

NON-LINEAR PEAK
ELIMINATION FILTER

Figure 40.   Difference Image Distributions, With and Without
the Non-Linear Peak Elimination Filter

158

5.4   Summary

In this chapter an adaptive Kalman filter was formulated to perform the reference image update function in a generalized image tracking system. The filter was analysed to determined its stability characteristics and its ability to significantly reduce the reference image noise variance was demonstrated through simulation. In addition, the use of the non-linear peak elimination filter as a prefilter for the sensor data was shown to not affect the difference image statistics other than to reduce the sample variance.

In the next chapter, the various pieces of a tracking system that have been developed in chapters three, four, and five will be integrated into a single algorithm and evaluated as a whole via simulation using real sensor data.

Chapter 6

TRACKER PERFORMANCE

Up to this point, the individual techniques for improving tracking performance have been analyzed and demonstrated in isolation. In this chapter, an integrated tracking algorithm is proposed which incorporates the concepts developed in previous chapters, and the performance of this integrated tracking algorithm is demonstrated in the presence of both noise and image change.

6.1   An Integrated Tracking Algorithm

An integrated tracking algorithm was developed to incorporate the nonlinear peak elimination prefilter, the adaptive, reference-set selection process using the gradient-magnitude estimation algorithm from Section 4.3.1 and the adaptive Kalman filter to perform the reference-image update function. The logic flow for this algorithm is shown in Figure 41.

When this algorithm was implemented for computer simulation the following features were included:

1) The number of pixels in the comparison set was adjustable up to a value of $N = 1024$ (limited by computer memory).

2) The data image source was selectable between either a sequence of noise corrupted copies of a fixed image or one of the three data image sequences discussed in Appendix I.

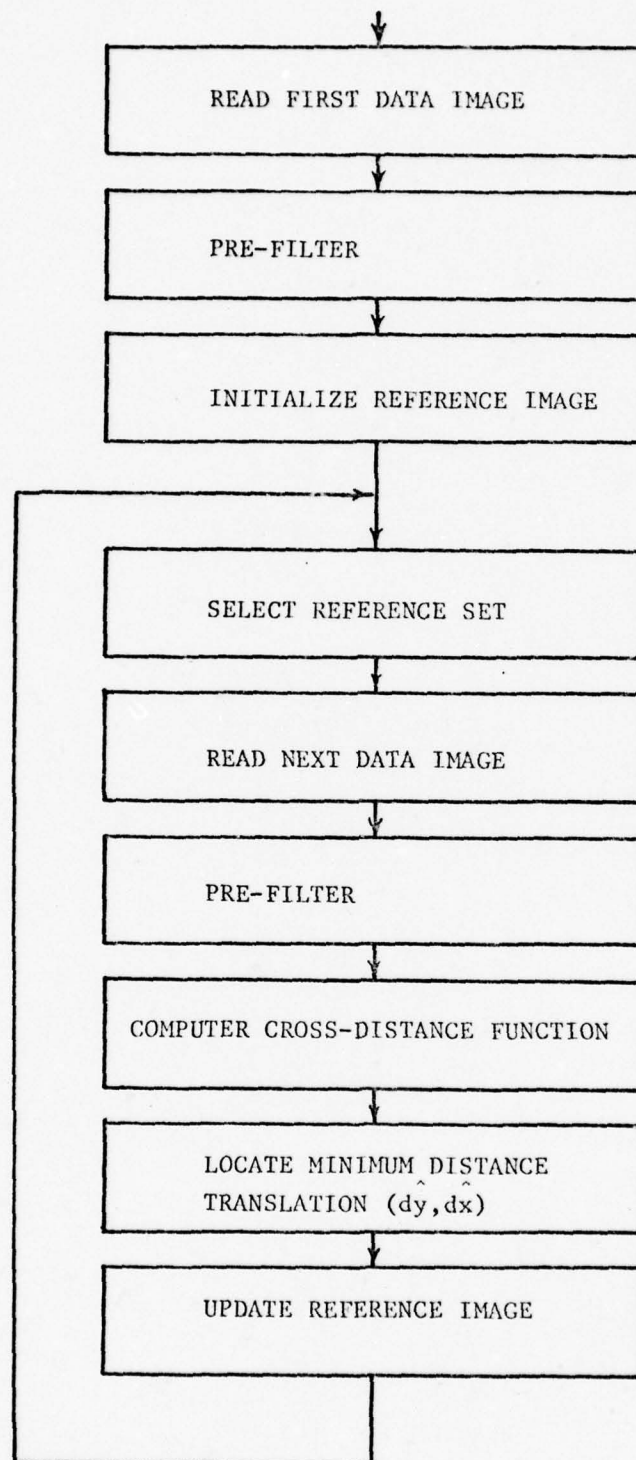3) Each of the three component algorithms could be turned off to allow the effects of its absence to be evaluated.

160

Figure 41.   Integrated Tracking Algorithm

161

When the nonlinear peak elimination prefilter was turned off, no prefiltering was performed. When the adaptive, reference-set selection algorithm was turned off, the reference-set was taken as a fixed grid of pixels centered on the aimpoint with a selectable inter-pixel spacing. Thus, a contiguous block of pixels could be used for one simulation run, and a sparse grid could be used for another. When the Kalman filter was turned off, the reference image update process simply copied the data image as the source for extracting the next reference set. This procedure presented the maximum number of opportunities for the tracker to accumulate error in a fixed amount of computer simulation time.

The performance of the tracker can be separated into two parts for evaluation purposes. The first part is the performance of the similarity detector (with or without prefiltering) in the presence of noise in the data image. In Section 3.10 it was shown that the sum of the noise in the reference image and the noise in the data image is the factor which determines probability of error and thus, for a particular image, the mean-square tracking error. The second part is the performance of the adaptive Kalman filter in estimating the underlying image from the data image sequence. While the filter can never reduce the reference-image noise component to zero, it can come very close to reducing the sum of $\sigma^2_{n,REF}$ and $\sigma^2_{n,DATA}$ by a factor of two from what it would be without the filter (without the filter,

$\sigma^2_{n,REF} = \sigma^2_{n,DATA}$). Recall that for moderate values of N, a signal-to-noise ratio improvement of a factor of two can make a very

162

significant contribution to reducing the probability of error (see Figure 11).

In Section 5.3 the performance of the adaptive Kalman filter was demonstrated with respect to its ability to reduce the reference image noise variance. That performance is independent of N and depends only on the ability of the similarity detector to provide a sequence of registered images. The similarity detector performance was measured by Monte Carlo simulation. Using the image in Figure 63 as the reference image, $N = 32$, 64, 128, and 1024, and the adaptive reference selection algorithm, 100 noisy data images were matched against the known perfect reference set. The mean-square registration error was computed for each set of 100 data images and is shown in Figure 42 for various values of $\sigma_{n,DATA}^2$. Monte Carlo runs of 64, 81, 100, and 200 images were made for $N=64$ with only small changes in mean-square error.

The nonlinear peak elimination filter provides approximately a 20% reduction in mean-square error for this particular image at a noise variance of 25. Figure 43 illustrates the decrease in average signal strength that accompanies the increase in N for this particular reference image for a shift of +1 pixel along the x-axis. This phenomenon of decreasing mean squared error in the face of decreasing average signal strength serves to illustrate the fact that increasing the number of elements in the reference set more than offsets the decrease in average signal strength.

Up to this point, all simulations have used a single known underlying image. While this technique provides excellent control over the simulation parameters and absolute knowledge about the relative
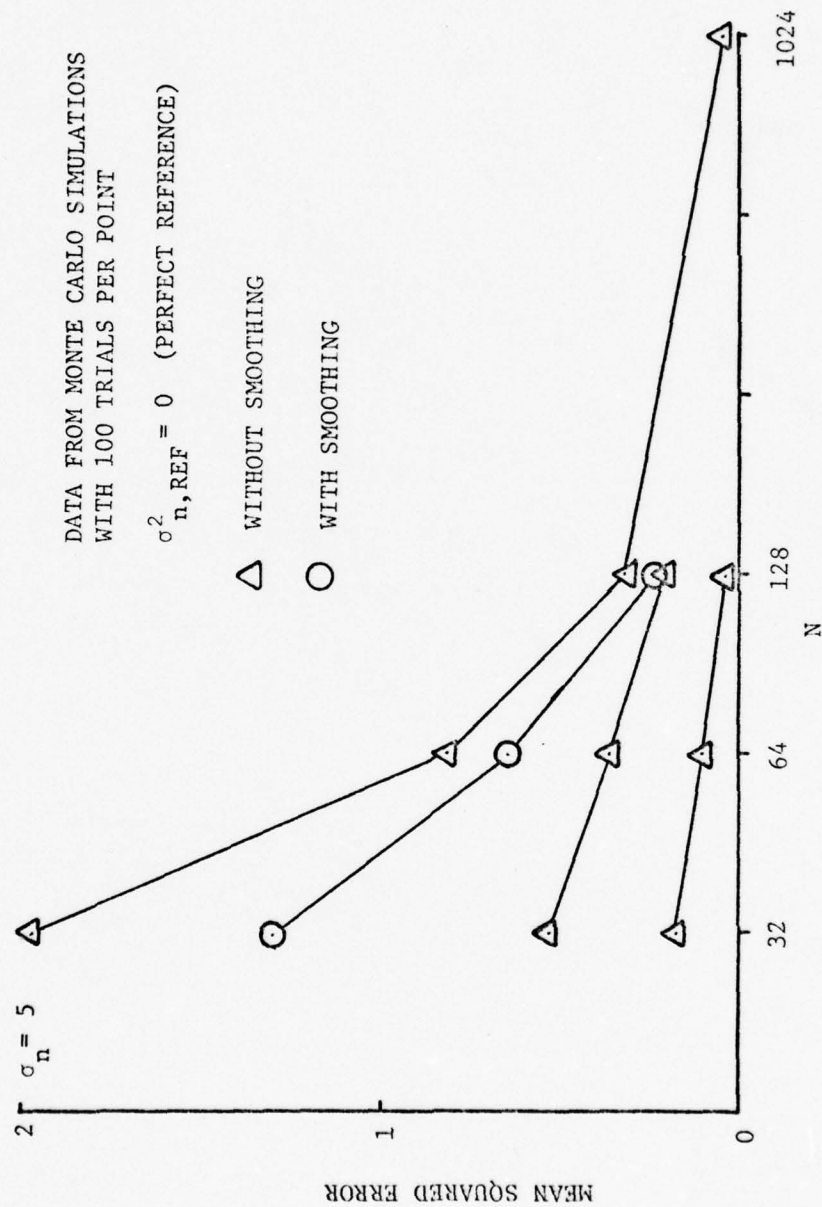
163

Figure 42. Detection Performance in the Presence of Noise
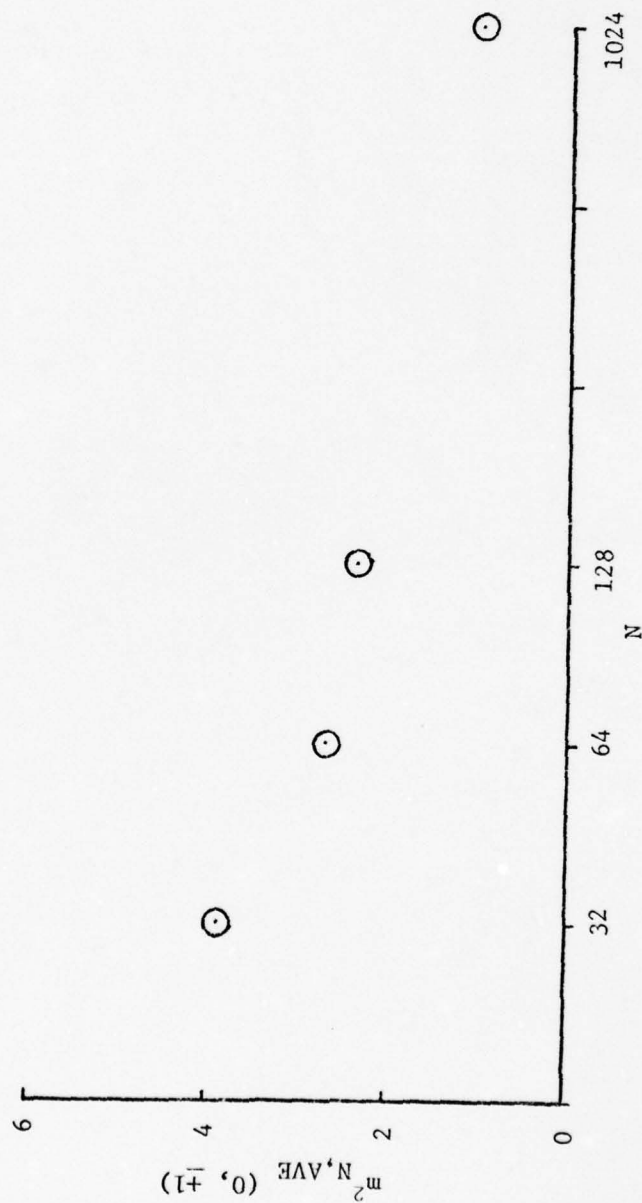
164

Figure 43. Average Signal Strength as a Function of Reference Set Size

motion of the image on a frame-to-frame basis, it does not allow exploration of the capability of the integrated tracking algorithm to track an object which is truly changing size and shape as well as position within the image. It is the potential ability of the adaptive Kalman filter to maintain an accurate estimate of the underlying image in the presence of change that promises to improve the performance of the total tracking system. The capability of adapting to a changing scene and deriving a measure of system performance from the noise variance estimates and the measured signal strength will provide a system designer with features not previously available in image-tracking systems.

As an example of the performance that can be obtained through integration of nonlinear prefiltering, adaptive reference selection and the adaptive Kalman Filter, the integrated tracker was used to track the data image sequence AIRPLANE for 89 frames. The nose of the aircraft was designated in the first frame, and the detected frame-to-frame motion was accumulated to produce the estimated position of the aircraft within each sequential image. After each reference image update, the reference image was written onto magnetic tape with the location of the estimated target position marked with a cross hair. A sample of these reference images with the indicated aimpoints marked is shown in Figure 44 through Figure 48.

These images correspond to the reference image after being updated from the corresponding image in Figure 58 through Figure 62. Figure 49 shows the detected motion of the image sequence with every tenth frame numbered (some frames had no detected motion so there are not necessarily nine points between each marked frame).
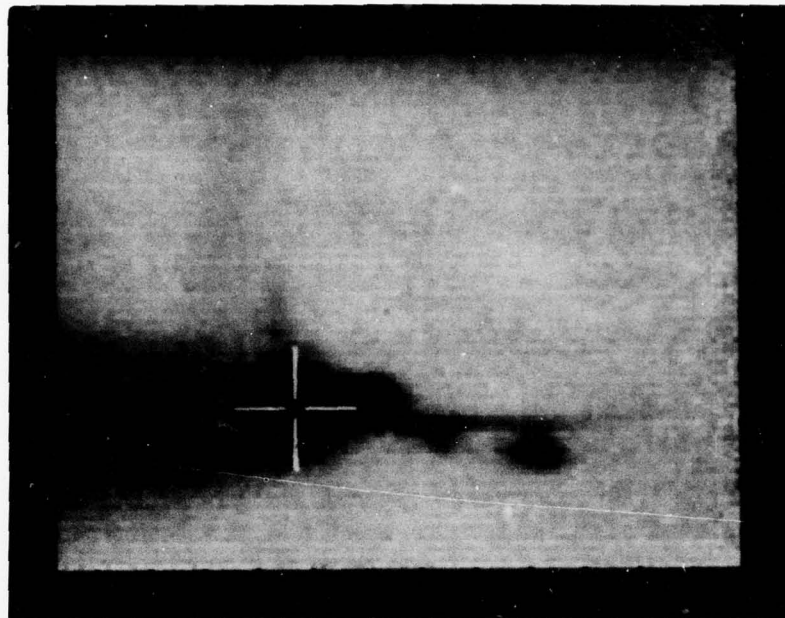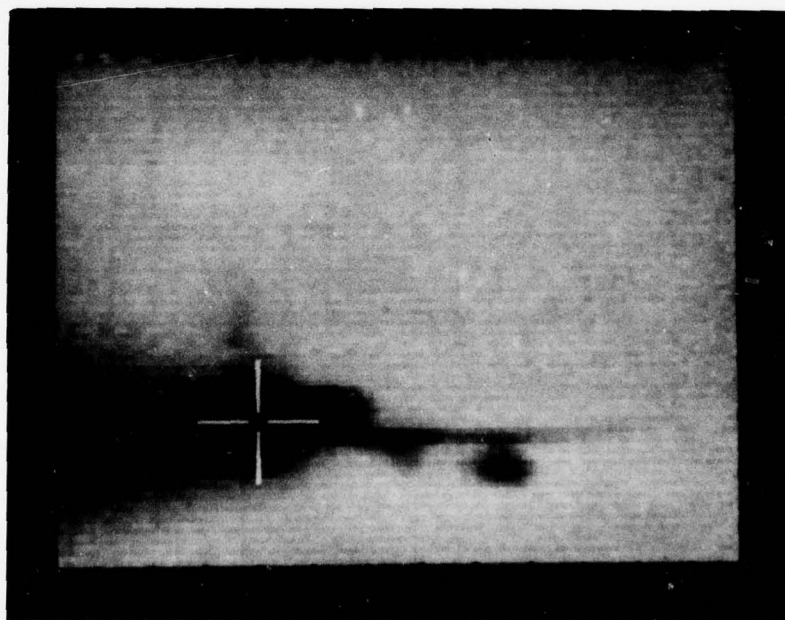
166
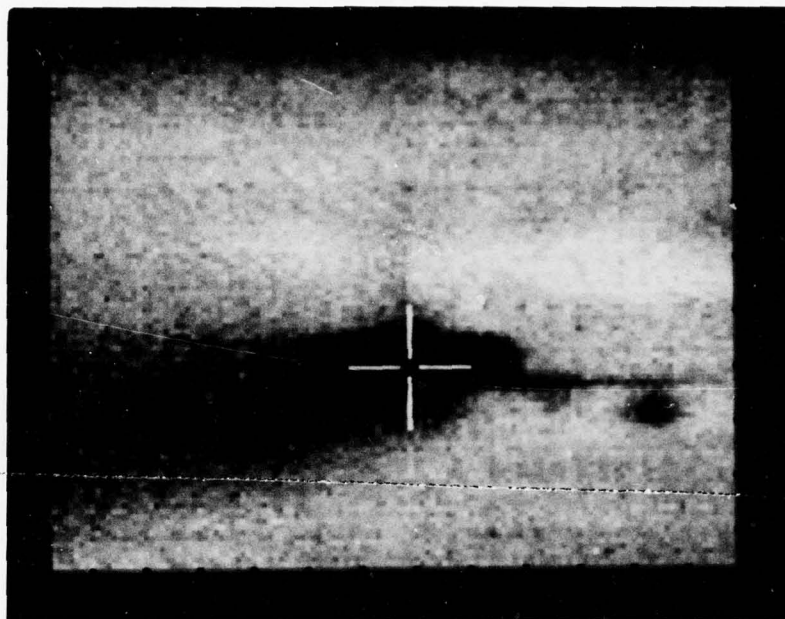
Frame 1

Frame 11

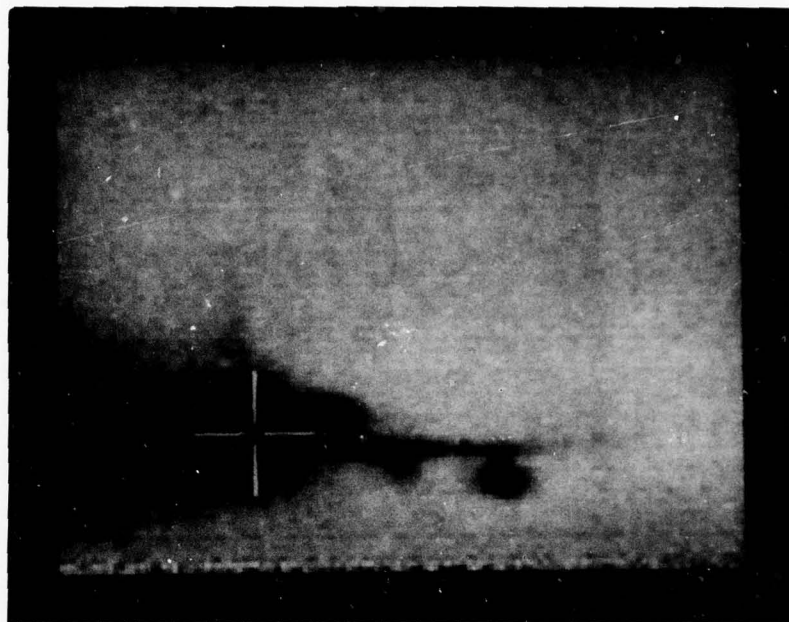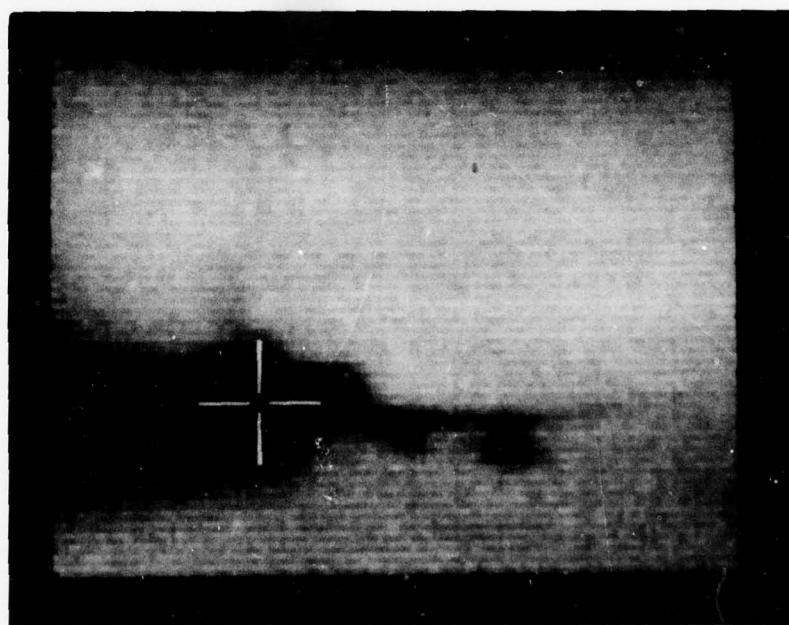Figure 44. Updated reference image after frames 1 and 11

167

Frame 21

Frame 31
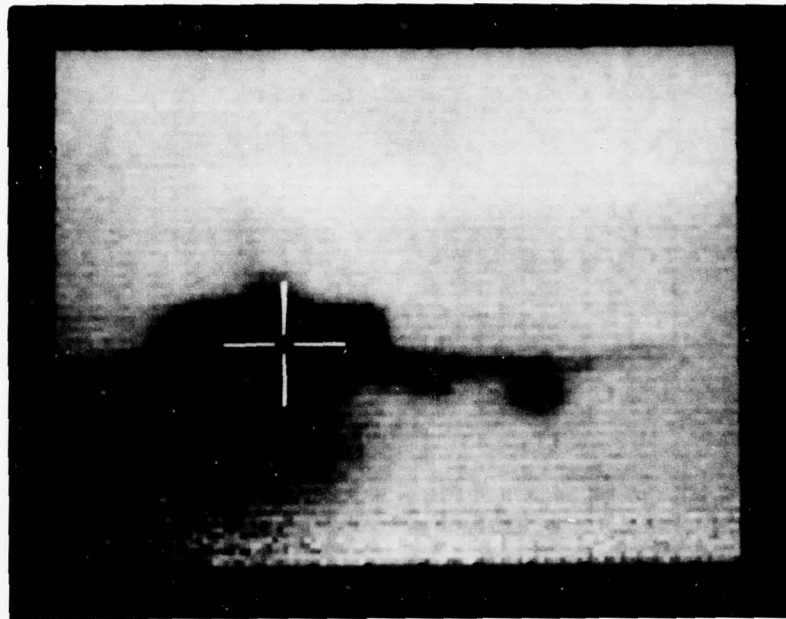
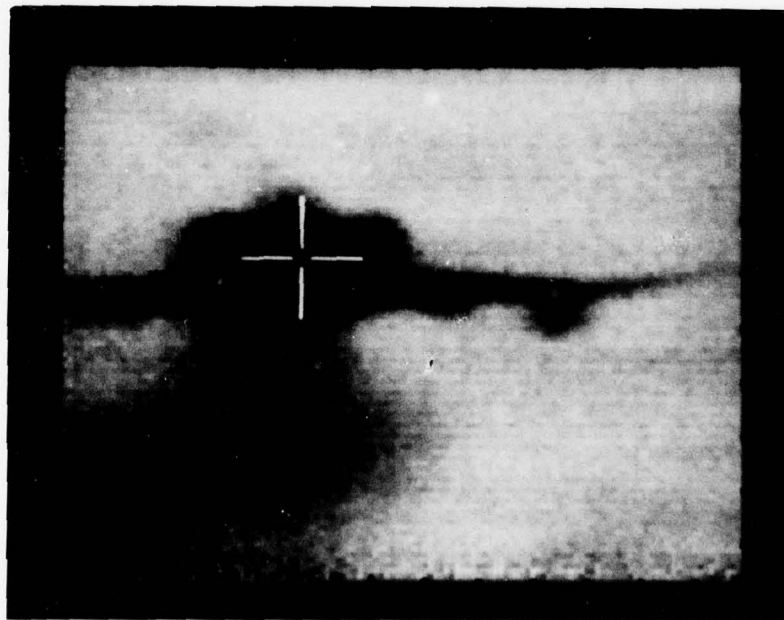Figure 45. Updated reference image after frames 21 and 31

Frame 41

Frame 51

Figure 46. Updated reference image after frames 41 and 51
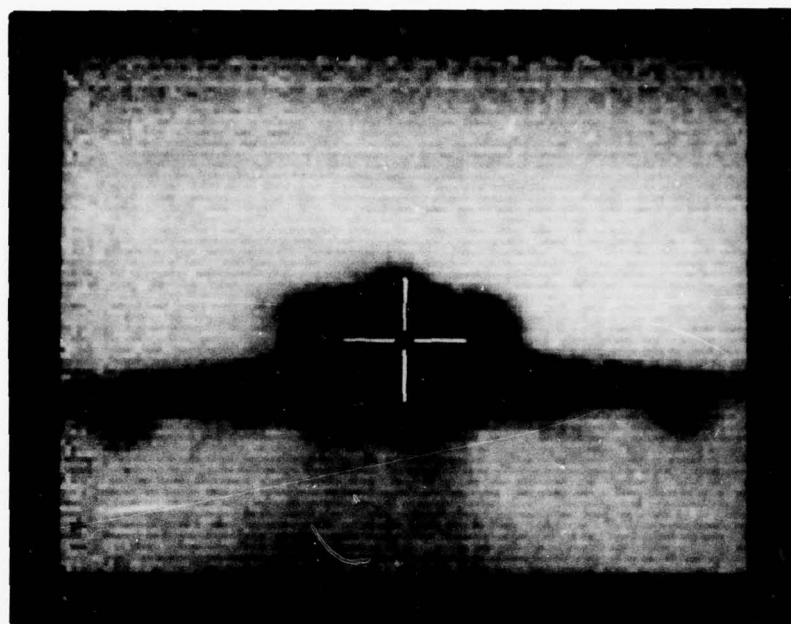
169

Frame 61

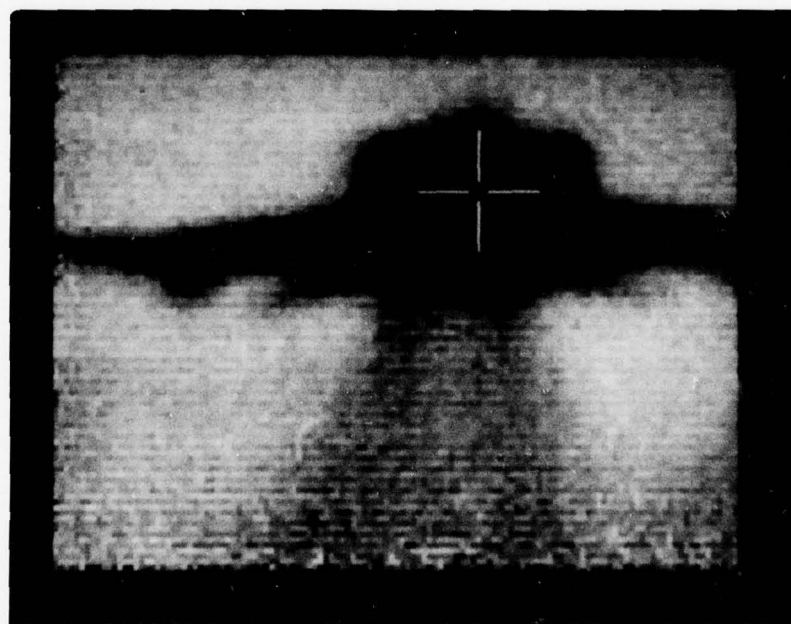Frame 71

Figure 47. Updated reference image after frames 61 and 71

170

Frame 81

Frame 89

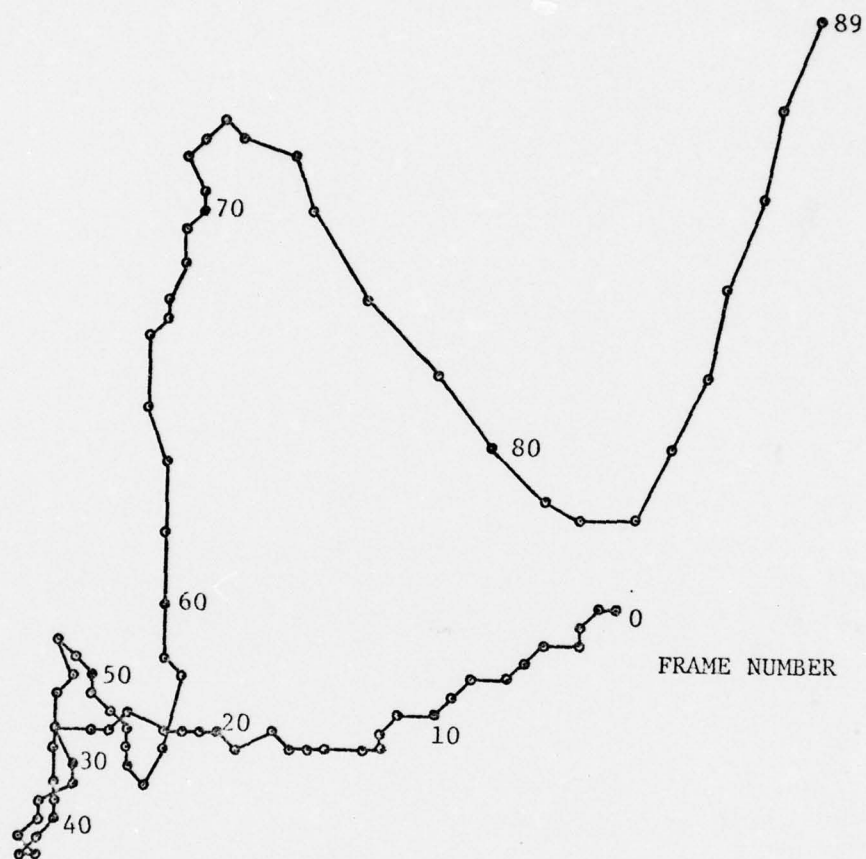Figure 48. Updated reference image after frames 81 and 89

171

Figure 49. Detected Image Motion for Image Sequence
AIRPLANE, N=128, with Non-Linear Prefilter
and Kalman Filtered Reference Update

The filter estimates of $\sigma^2_{n,DATA}$ and $\sigma^2_{n,REF}$ are shown in Figure 50 and the variance of the difference image is plotted in Figure 51.

For this simulation, there were 128 elements in the reference set. For comparison, Figure 52 shows the detected motion for a minimum norm tracker using a 256-pixel reference set arranged in a contiguous block (16 by 16) centered at the initial position indicated by the cross hair in Figure 44 but with no prefiltering and no Kalman filtered reference update.

While this tracker takes maximum advantage of the correlation between correct and incorrect trial registrations, and uses twice as many pixels in the reference set, it cannot track the motion of the image sequence.

Figure 53, Figure 54 and Figure 55 illustrate the performance of the integrated tracking algorithm in the presence of added noise.

Two important tracker characteristics are demonstrated in this tracking sequence. First, the filter was initialized with a value of $\sigma^2_{n,DATA}(0) = 29$ which is an unnecessarily pessimistic value. The filter however rapidly diagnosed that this value was not consistent with the observed difference image sample variances and reduced the estimate of $\sigma^2_{n,DATA}$ to about 20 over a period of 30 frames (2 time constants for the selected value of β). This value then remained approximately constant for the rest of this run with the exception of a perturbation around frame 65 due to a loss of track and the ensuing
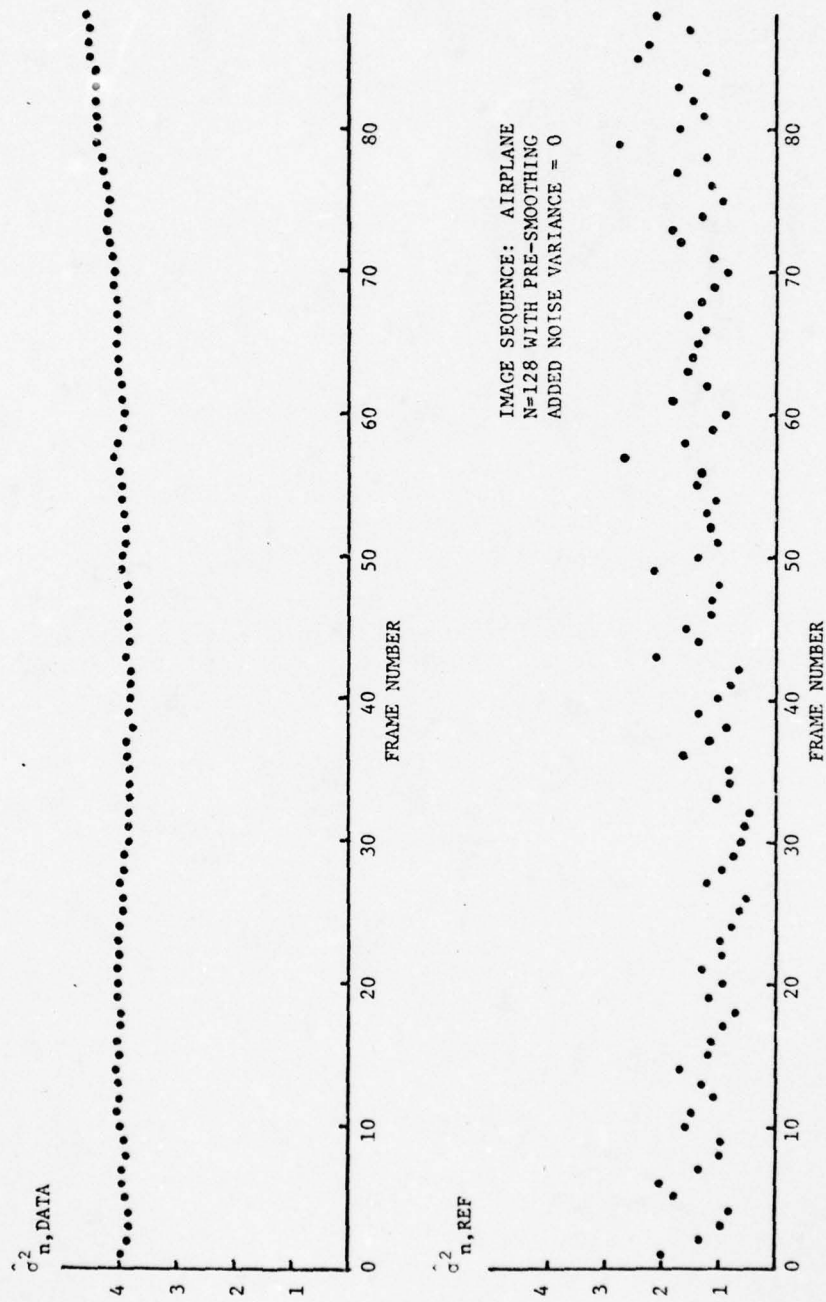
173

Figure 50. Kalman Filter Variance Estimates for Integrated Tracking
Algorithm (Image Sequence AIRPLANE)

IMAGE SEQUENCE: AIRPLANE
N=128 WITH PRE-SMOOTHING
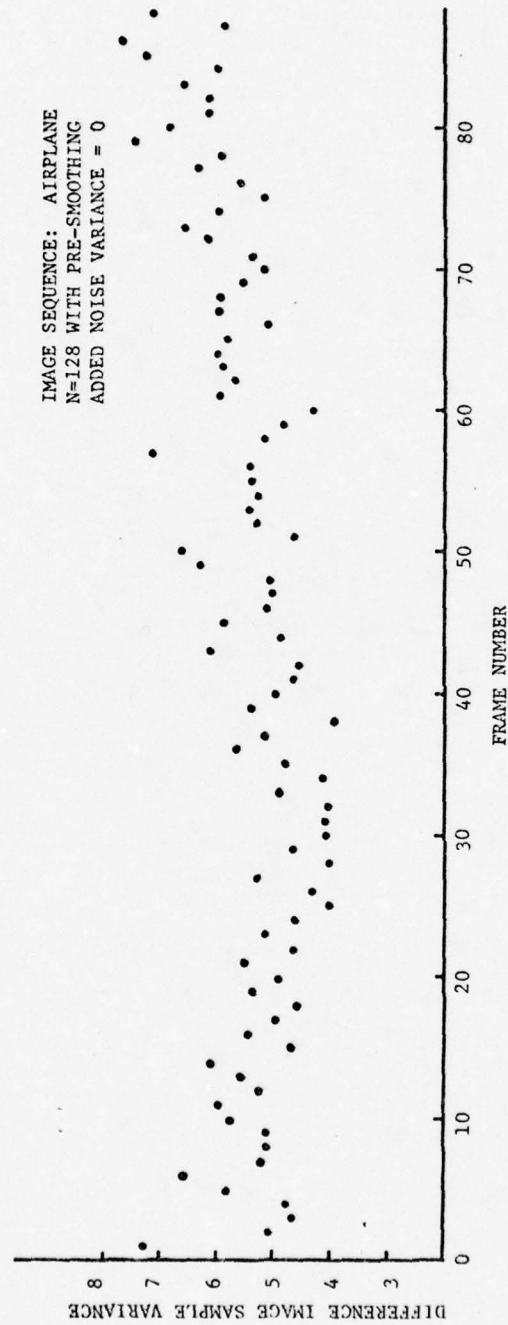ADDED NOISE VARIANCE = 0

Figure 51. Difference Image Sample Variance for Integrated Tracking
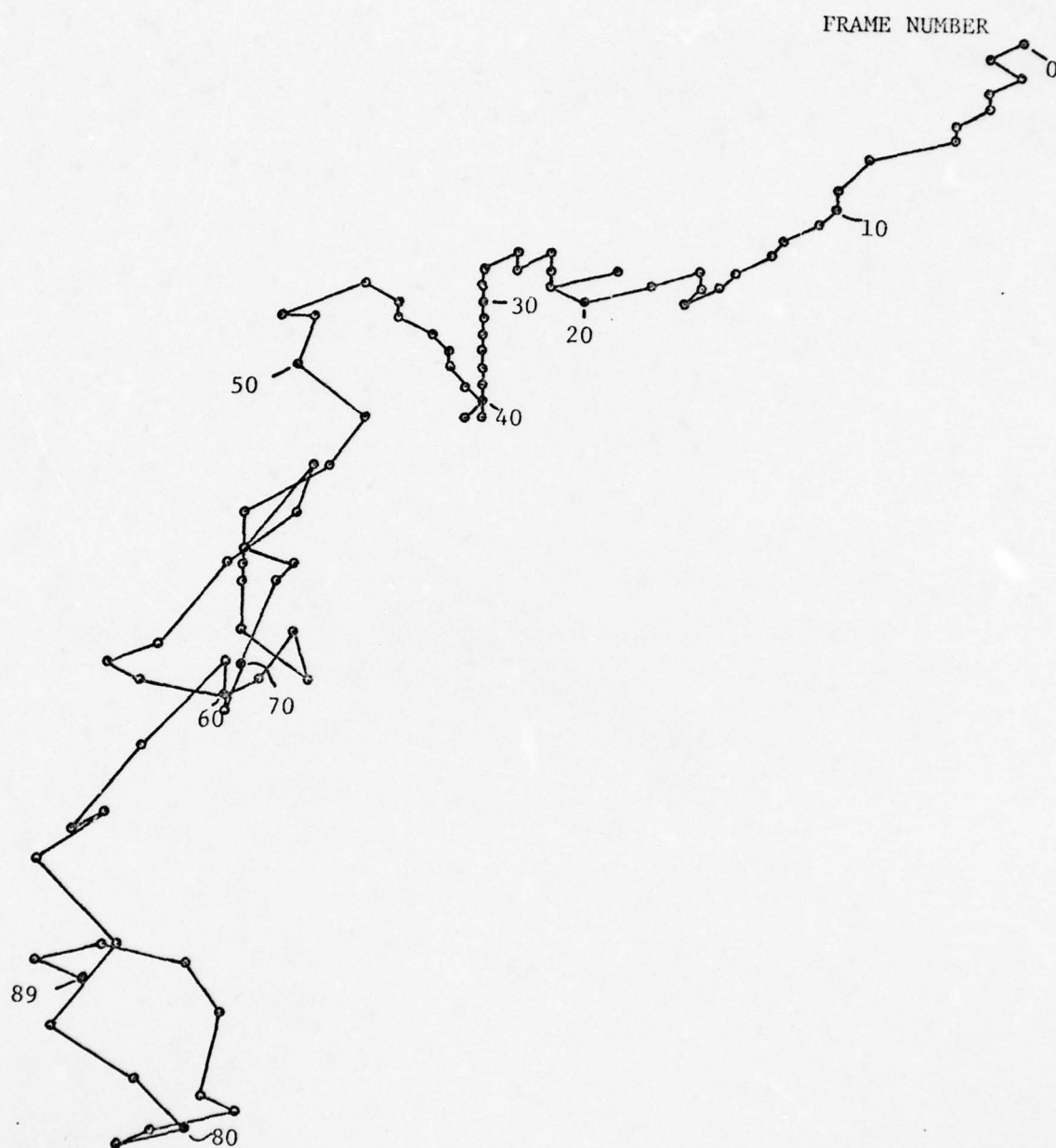Algorithm (Image Sequence AIRPLANE)

175

Figure 52. Detected Image Motion for Image Sequence AIRPLANE Using a 16x16 Block Reference and No Prefilter or Reference Update Filter

176
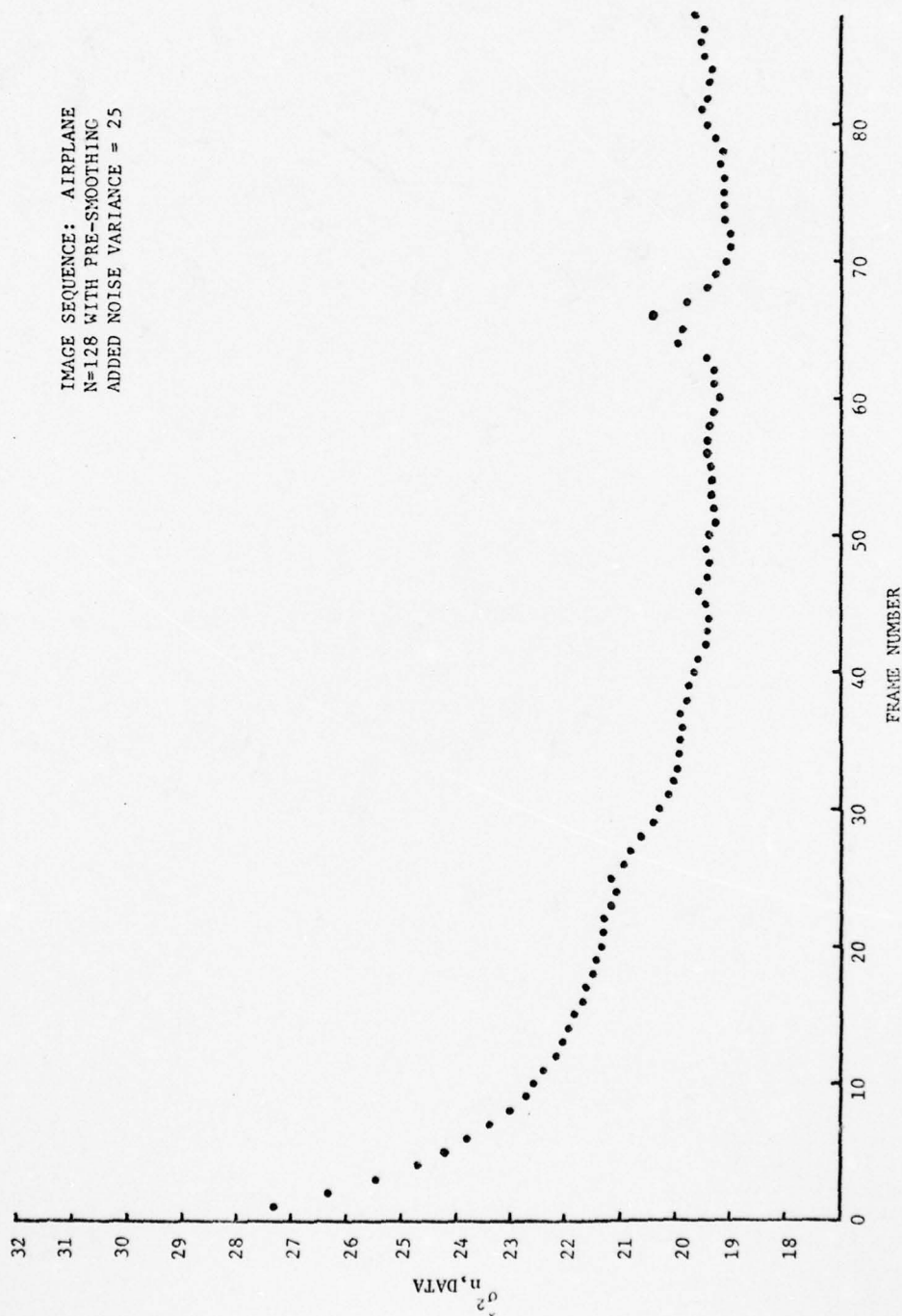
Figure 53. Sensor Noise Variance Estimate for Integrated Tracking Algorithm when Tracking Image Sequence AIRPLANE in the Presence of Added Noise

IMAGE SEQUENCE: AIRPLANE
N=128 WITH PRE-SMOOTHING
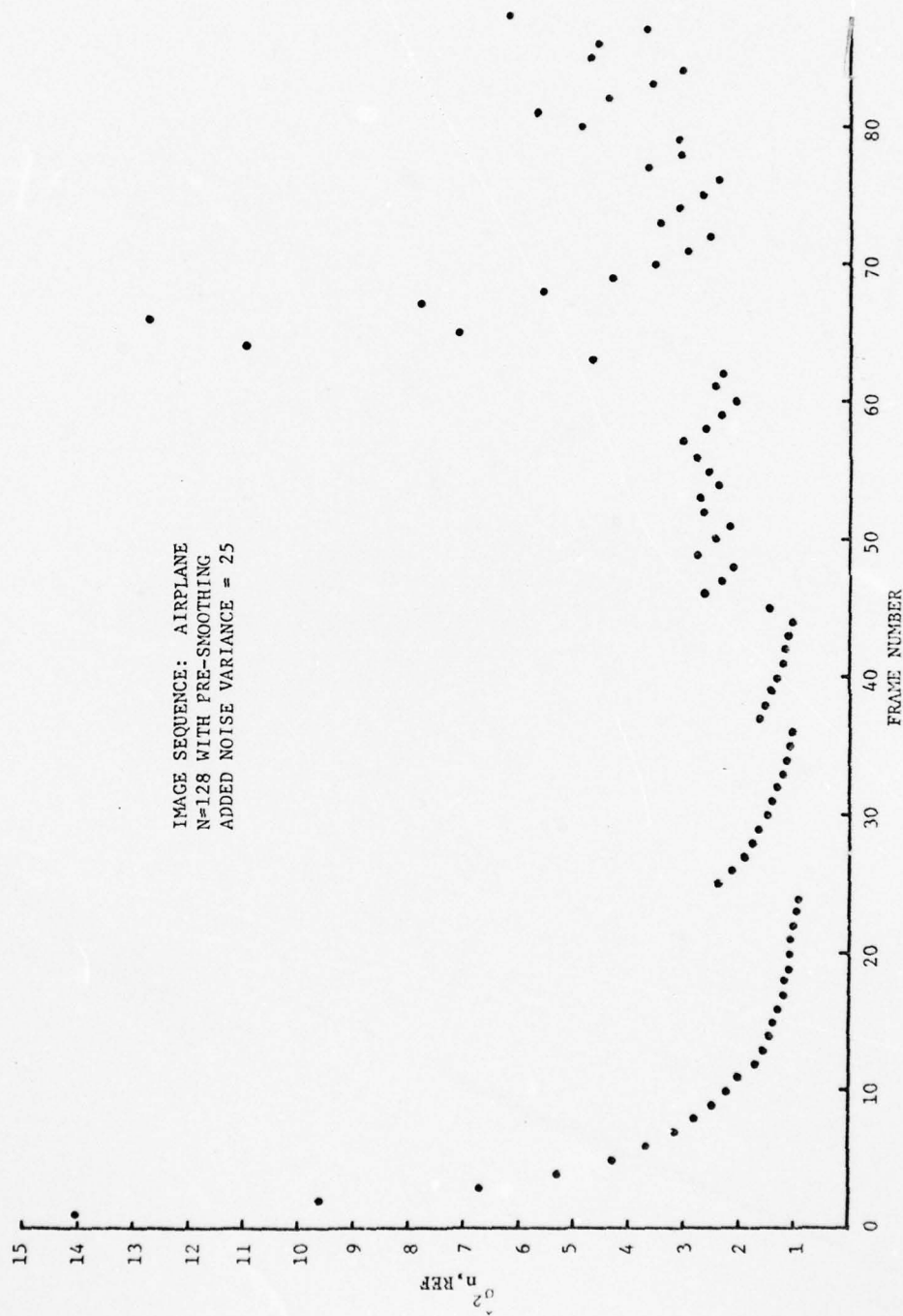ADDED NOISE VARIANCE = 25

Figure 54. Reference Image Error Variance Estimate for Integrated Tracking Algorithm when Tracking Image Sequence AIRPLANE in the Presence of Added Noise
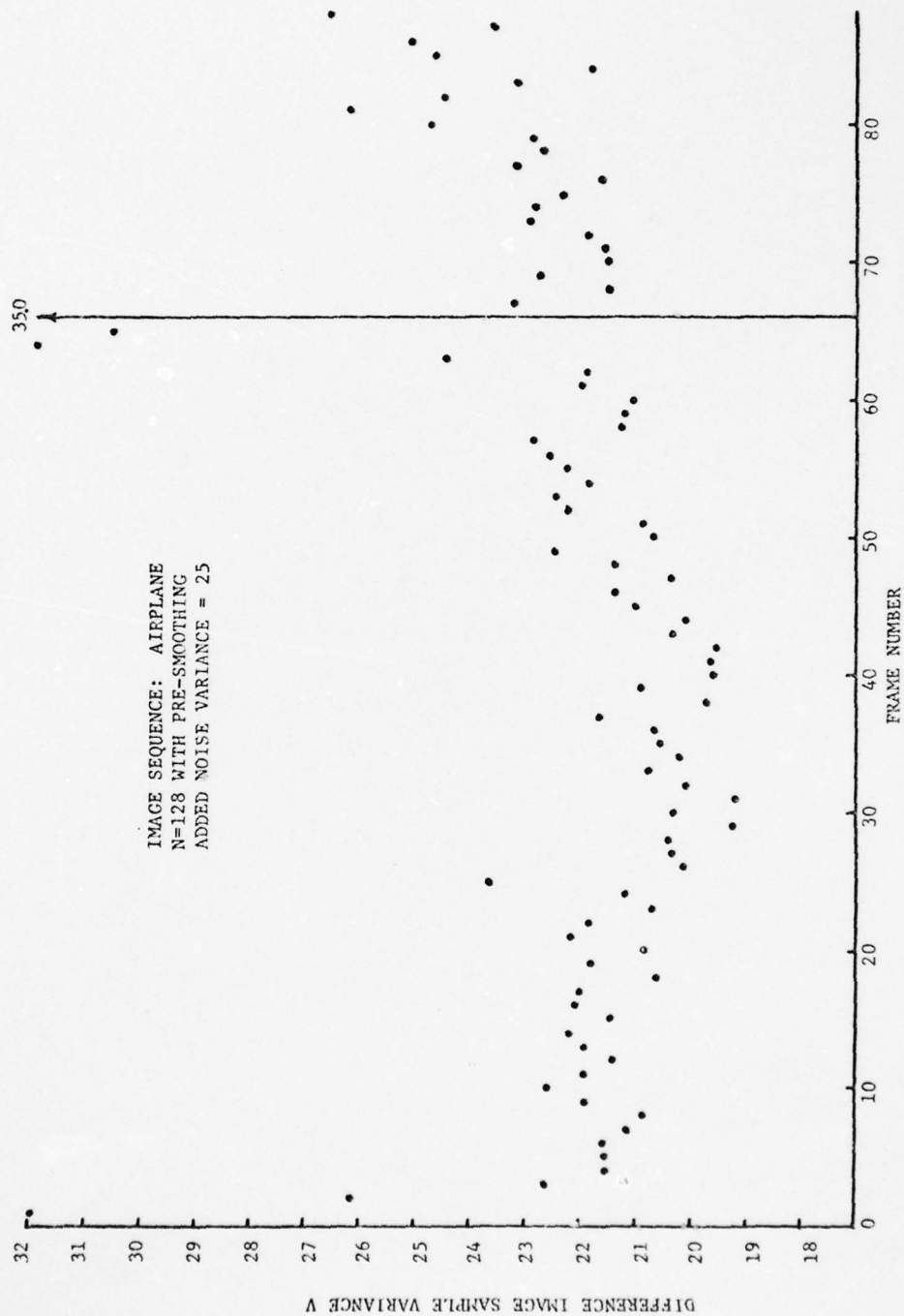
178

Figure 55. Difference Image Sample Variance for Integrated Tracking Algorithm when Tracking Image Sequence AIRPLANE in the Presence of Added Noise
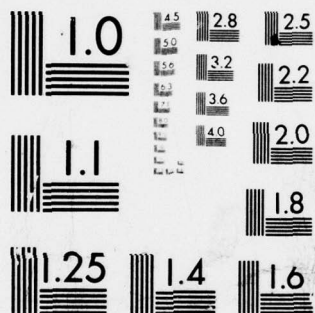
179
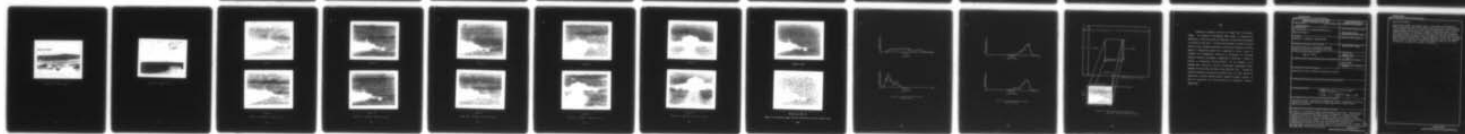
MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

reacquisition. Second, the loss of track which occurs at frame 63 is immediately recognized by the filter as a significant event. The very rapid increase in $\hat{\sigma}^2_{n,REF}$ which allows the reference image to change quickly, is a direct result of the filter entering Region III of operation (the unstable region). By frame 67 the filter has reentered Region II and has reduced $\hat{\sigma}^2_{n,REF}$ to near its previous value by frame 71.

The gradual increase in $\hat{\sigma}^2_{n,REF}$ from frame 72 to the end of the data sequence is attributable to the variation in size and shape of the aircraft image, and demonstrates the ability of the filter to accommodate itself to actual image change.

The conclusions to be drawn from these simulations and those of Section 5.3 are that:

1) Adaptive reference selection maximizes the signal component of the reference set.

2) Performance of the minimum norm tracking algorithm as measured by mean square registration error improves with increasing N and also improves with increasing signal-to-noise ratio for fixed N.

3) The nonlinear peak elimination prefilter reduces mean square tracking error through reduction of the data image noise variance.

4) The adaptive Kalman filter can significantly reduce the reference image noise variance and simultaneously estimate both the reference image noise variance and the data image noise variance.

180

6.2   Summary

In this chapter the performance characteristics of an
integrated tracking algorithm have been investigated. The selected
algorithm incorporated the non-linear peak elimination prefilter,
adaptive reference set selection using the gradient magnitude histogram
for selecting reference pixels, and the adaptive Kalman filter for
reference image update. The benefits of increasing reference set size
were illustrated and the capability of the integrated tracking
algorithm to accurately track an image sequence was demonstrated.

Chapter 7

CONCLUSIONS AND RECOMMENDATIONS

The major thrust of this research was directed toward
developing new techniques for tracking sequences of digitized images.
A model of a generalized image tracking system was defined for use as a
basis for analysis, and four new techniques were developed. The
practical implications of these techniques are summarized in the next
section, as are the conclusions which can be drawn from this work. In
the last section several recommended areas for future research are
pointed out.

7.1   Summary and Conclusions

Four new techniques were developed for application to the
general sequential image tracking problem:

    1) A non-linear peak elimination prefilter

    2) Two techniques for similarity detection:

        a) A non-uniformly weighted norm

        b) An adaptive reference set
selection algorithm based on the
gradient magnitude histogram (including
a new and very effective gradient
magnitude estimator)

    3) An adaptive Kalman filter to perform the
reference image update

While the four techniques which were developed are applicable
to three different functional areas in the general image tracking
system, and with the exception of the non-uniformly weighted norm and

the adaptive reference set selection algorithm which are not directly comparable, it is possible to provide a subjective evaluation of their relative merit.

The greatest payoff is obtained by using the adaptive Kalman filter to maintain a high accuracy, low noise reference image at all times. The effective signal-to-noise ratio for the tracker is almost doubled when the filter is used, with a corresponding improvement in tracker performance.

The next most usefull of the techniques developed is the adaptive reference set selection algorithm. The performance improvement which is obtained by tracking on subsets of the reference and data images comes from the correspondingly larger image that can be processed. For example, a tracker which today can process 256 reference pixels and 256 trial registrations per frame may only maintain a data image containing 1024 total pixels and a reference image of 256 pixels. By using the adaptive reference set selection algorithm, a much larger reference image can be maintained (perhaps as large as the entire data image) while only processing a small subset to determine image misregistration. The resulting signal-to-noise ratio is substantially enhanced by using only the "good" pixels for the reference set and the same processor speed can be tolerated. The one factor on which this projection depends is the availability of a device to perform the gradient estimation task at realtime rates. While the gradient estimator developed in Chapter 4 has many attractive features, a less complex gradient estimator in hardware might prove to be satisfactory in implementing the adaptive reference set selection algorithm for a particular application.

183

The nonlinear peak elimination prefilter appears to be very easy to implement in either hardware or software, and for low contrast images seems to provide up to a factor of two reduction in noise variance. However, for high contrast imagery the adaptive reference set selection algorithm will incorporate into the reference set pixels which lie on high gradient edges, and are less likely to have been affected by the prefilter. Under this condition, the nonlinear peak elimination filter may not provide its maximum potential benefit.

An important aspect of the integrated tracking algorithm is the serendipitous behavior of the component parts. The non-linear peak elimination prefilter reduces the random noise component of the incoming data images. The adaptive reference set selection algorithm maximizes the signal component of the reference set so that the minimum distance registration is correct a higher percentage of the time, thus reducing the average difference image sample variance. The adaptive Kalman filter maintains a high quality (low noise) reference image and estimates both the data image noise variance and the reference image noise variance. The gradient magnitude estimator uses the Kalman filter estimate of the reference image noise variance to control the detection threshold and thus maintains a fixed probability of erroneously including a bad pixel in the reference set. Since the pixels in the reference set tend to lie along edges in the image, the natural adjacency of the reference set pixels takes advantage of the noise correlation that exists between the correct and incorrect trial registrations and reduces the probability of selecting an incorrect trial registration. The reduced noise component of the Kalman filtered reference image further decreases the probability of error.

184

7.2   Recommended Future Work

Several areas of potentially fruitful research are apparent. Since the smoothness of a two dimensional distance function determines the appropriateness of the search type, an exhaustive search is dictated where the trend information from adjacent trial registrations is an unreliable indicator of the direction toward the minimum of the distance function. It seems reasonable to expect that a roughness parameter can be developed which is a function of the signal-to-noise ratio. This parameter should indicate the probability that the minimum distance lies in the direction indicated by a distance function gradient measure associated with a particular trial registration.

The development of techniques to determine the location of the registration coordinates by interpolating between trial registrations would be a useful extension of this research. This seems to have some potential for reducing error.

The question of when to extract a new reference set from the reference image in order to minimize the aimpoint drift rate remains unanswered, as well as a number of questions regarding the relative performance of trackers employing more easily computed distance functions (absolute value or Hamming distances for example).

In the near future there is probably a speed advantage to be had in any digital processor using fixed point arithmetic. As a result there are questions to be answered regarding the appropriate word length and scaling to be used in mechanizing the adaptive Kalman filter.

References

1. Anuta, Paul E. "Spacial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques." IEEE Transactions on Geoscience Electronics, pp353-368, October 1970.

2. Arcese,A.,P. H. Mangert, and E. W. Tromboni. "Image Detection Through Bipolar Correlation." IEEE Transactions on Information Theory, Vol IT-16, pp. 534-541, (Sept 1970).

3. "Visual Tracking Unit Aids Pilot", Aviation Week and Space Technology, (18 November 1968), p99.

4. "Problems Balk CONDOR Production Plans", Aviation Week and Space Technology, (17 February 1969), pp18-19.

5. "MAVERICK Enters Flight Test", Aviation Week and Space Technology, (5 January 1970), p22.

6. "AIM-82 Missile Proposals for USAF F-15 Fighter Readied",Aviation Week and Space Technology, (2 March 1970), p19.

7. Backensto, W. V., and M. Y. Pines. Low Cost Arrays for Detection of Infrared. Air Force Avionics Laboratory Report, AFAL-TR-77-65, December 1976.

8. Barnea, Daniel I., and Harvey F. Silverman. "A Class of Algorithms for Fast Digital Image Registration." IEEE Transactions on Computers, Vol C-21, (February 1972), pp. 179-186.

9. Bernstein, R., and H. Silverman. "Digital Techniques for Earth Resource Image Data Processing." AIAA Paper 71-978, Presented at the 8th Annual AIAA Meeting, October 1971.

10. Blum,Marvin. A Method of Tracking an Interior Point of a Plane Figure on a Target Imaging Device. Northrop Corporation, NRTC-74-9R, April 1974.

11. Dobbins, Harry M., and Michael J. Noviskey. A High Data Rate Digital Spacial Frequency Tracking System. Air Force Avionics Laboratory Report, AFAL-TR-74-293, December 1974.

12. Gelb, Arthur, et al. Applied Optimal Estimation. Cambridge, Massachusetts: The MIT Press, 1974.

13. Hayes, Arthur F., Gary F. Braley, and Nelson R. Zagalsky. Estimation of Correlation and Decorrelation Functions of Imagery. Honeywell Inc. Report 12535-FRI, June 1967. (AD818333)

14. Highleyman, W. H., "The Design and Analysis of Pattern Recognition experiments." The Bell System Technical Journal, March 1962, pp 723-744.

15. Jenkins, G. M. and D. G. Watts. Spectral Analysis and its Applications. San Francisco: Holden-Day, 1968.

16. Johnson, Norman L., and Samuel Kotz. Continuous Univariate Distributions - 2. Boston: Houghton Mifflin Co., 1970.

17. Kuglin, C. D., and D. C. Hines. "The Phase Correlation Image Alignment Method." Unpublished paper.

18. Leese, J. A., C. S. Novak, and B. B. Clark. "An Automated Technique for Obtaining Cloud Motion from Geosynchronous Satellite Data using Cross-Correlation." Journal of Applied Meteorology, Vol 10, (February 1971), pp. 118-132.

19. Lillestrand, Robert L. "Techniques for Change Detection." IEEE Transactions on Computers, Vol C-21, (July 1972), pp. 654-659.

20. Melbourne William G. "Navigation Between the Planets", Scientific American, (June 1976), pp58-74

21. Miller, Barry. "Devices Gain Growing Weapons Role", Aviation Week and Space Technology, (19 January 1970), pp54-65.

22. -----. "New Roles Grow for Electro-Optics". Aviation Week and Space Technology, (22 June 1970), p155-167.

23. -----. "Weapon Delivery Aid Flexibility Sought", Aviation Week and Space Technology, (10 May 1971), pp75-82.

187

24. Nering, Evar D. Linear Algebra and Matrix Theory. New York: John Wiley and Sons, 1970.

25. Merha, Raman K. "Approaches to Adaptive Filtering." IEEE Transactions on Automatic Control, October 1972, pp693-698.

26. Papoulis, Athanasios. Probability, Random Variables and Stochastic Processes. New York: McGraw-Hill, 1965.

27. Parzen, Emanuel. Stochastic Processes. San Francisco: Holden-Day, Inc., 1962.

28. Paulson, David C. Autonomous Navigation Technology Phase 0 Study. Air Force Space and Missile Systems Organization Report, SAMSO-TR-72-93, May 1972.

29. Pratt, William K. "Correlation Techniques of Image Registration." IEEE Transactions on Aerospace and Electronic Systems, Vol AES-10, (3 May 1974), pp. 353-358.

30. Rosenfeld, A. "Automatic Detection of Changes in Reconnaissance Data." Proceedings of the 5th Conference on Military Electronics, 1961, pp. 492-499.

31. -----. Picture Processing by Computer. New York: Academic Press, 1969.

32. Shapiro, Gerald N. High Speed Micro Signal Processor Study. Air Force Avionics Laboratory Report, AFAL-TR-77-52, September 1977.

33. Silverthorn, J. Aerial Gunnery Firecontrol Algorithms. Air Force Avionics Laboratory Report. AFAL-TM-77-43-RW1-2, August 1977.

34. Smith, Eric A., and Dennis R. Phillips. "Automated Cloud Tracking Using Precisely Alligned Digital ATS Pictures." IEEE Transactions on Computers, (July 1972), pp. 715-729.

35. Taylor, Angus E. Introduction to Functional Analysis. New York: John Wiley and Sons, Inc., 1958.

36. Tremblay, J. P., and R Manohar. Discrete Mathematical Structures

with Applications to Computer Science. New York: McGraw-Hill, 1975.

37. Walker, W. S. et. al. Target Homing Study. Air Force Avionics Laboratory Report, AFAL-TR-65-312, November 1965. (AD 371368)

38. Webber, W. F., et al. Mathematical Research in Target Detection and Identification. McDonnell Douglass Astronautics Company, MDC G5105, February 1974. (AD777203)

39. Woody, William R. Project engineer for focal plane processing technology, Technology Division, Air Force Avionics Laboratory (personal interview). Wright-Patterson Air Force Base, Ohio, September 1, 1978.

40. Woolfson, Martin G. "Digital Area Correlation Tracker".Proceedings of the IEEE 1978 National Aerospace and Electronics Conference, (May 1978), pp882-886.

41. Woolfson, M. G., et al. Digital Correlation Tracker, Phase I Computer Simulation. Air Force Weapons Laboratory Report. AFWL-TR-74-170, March 1975.

Appendix I

Data Characteristics

Three image sequences were gathered for use in tracking experiments. Sample images from each sequence are illustrated in Figure 56 through Figure 62.

All image sequences were obtained from vidicon sensors, recorded on commercial video recorders, and transferred to a video disc for digitizing. A subsection 128 pixels wide by 96 pixels high from alternate T.V. fields was then digitized to six-bit accuracy with an inter-pixel separation of 100 nanoseconds in the scan direction. Total system video bandwidth was approximately 3.5 MHz. No correction was made for nonlinearity of the vidicon input-output transfer function. For purposes of reproduction the dynamic range of the images was expanded linearly to the point where one tenth of one percent of the brightest and darkest pixels were clipped. To obtain the original aspect for the images, tilt the page until the ratio of height to width is 1.17. Figure 64, Figure 65, and Figure 66 present the distribution of intensity levels for the first frame of each sequence.

Figure 67 shows the relationship between the original T.V. format and the images as reproduced here.

Both CARS and TREES were obtained under controlled conditions with a rigidly mounted, high quality, commercial T.V. camera. The sequence CARS presents a highway intersection with cars stopped at a red light. The image sequence TREES presents a field of ripe winter wheat containing some weed growth with a line of trees in the background.

190

AIRPLANE was obtained under less controlled conditions from a small ruggedized T.V. camera mounted on the gimbal of a ground-based aircraft tracking system. This image sequence presents an Air Force F-4 aircraft making a low pass over the tracking site. The aircraft is moving with respect to the background, the sensor field of view is moving with respect to the line-of-sight to the aircraft, and the aircraft is changing in both aspect and apparent size during the approximately three seconds represented by this image sequence.

Figure 56. CARS – frame 1
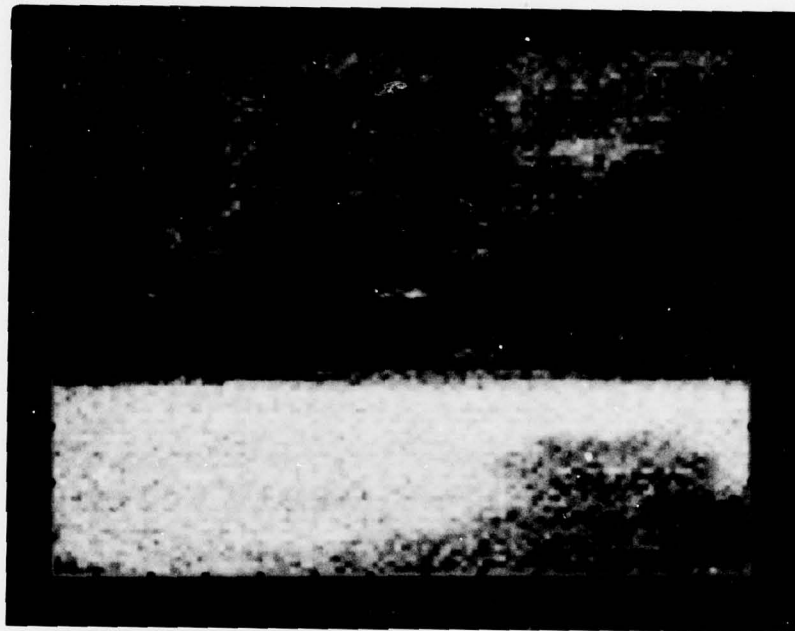
Figure 57. TREES - frame 1
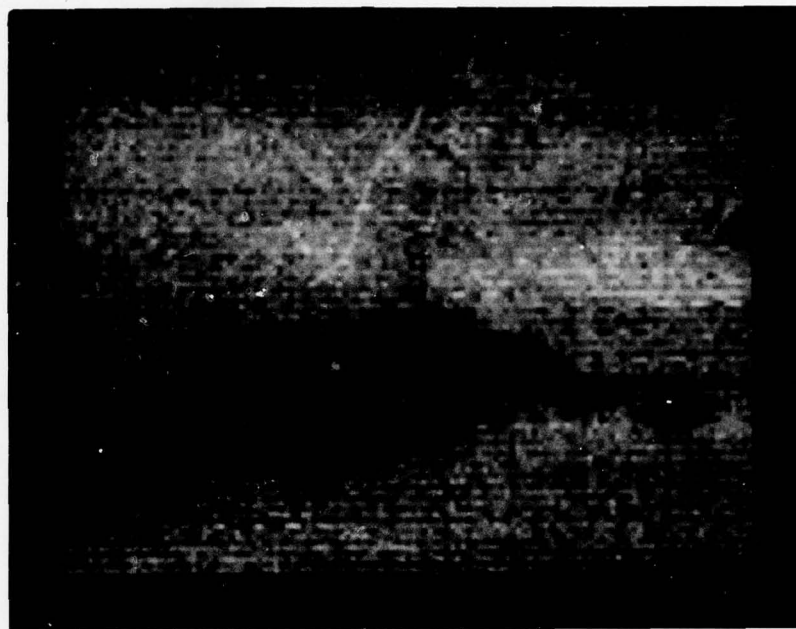
Frame 1
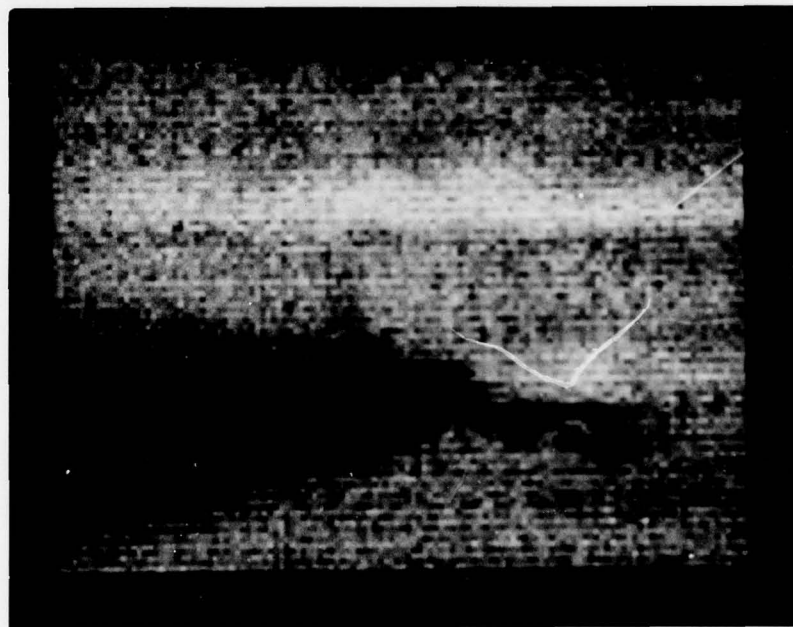
Frame 11

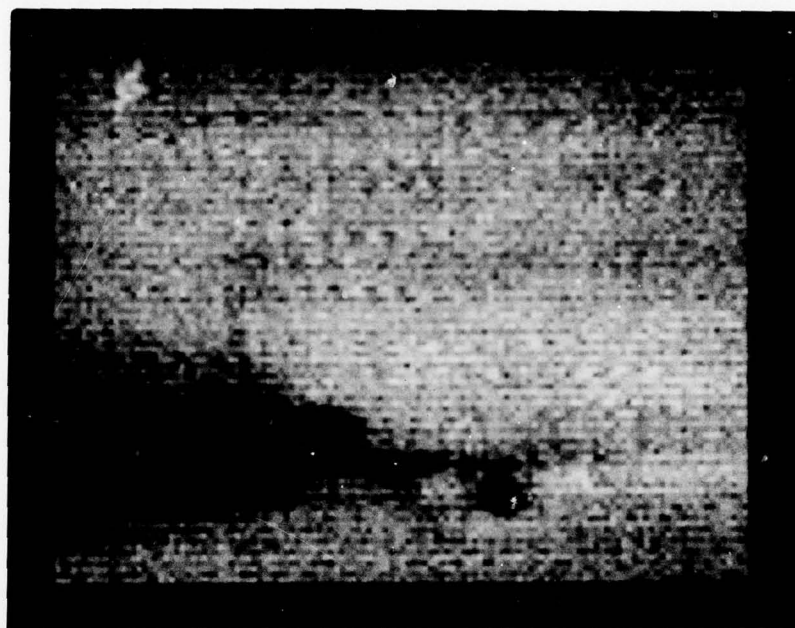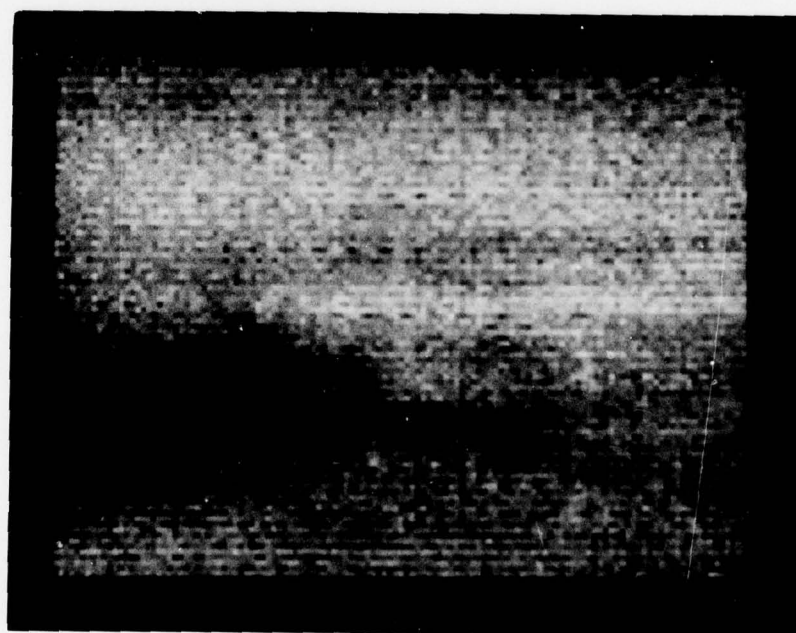Figure 58. AIRPLANE – frames 1 and 11

194
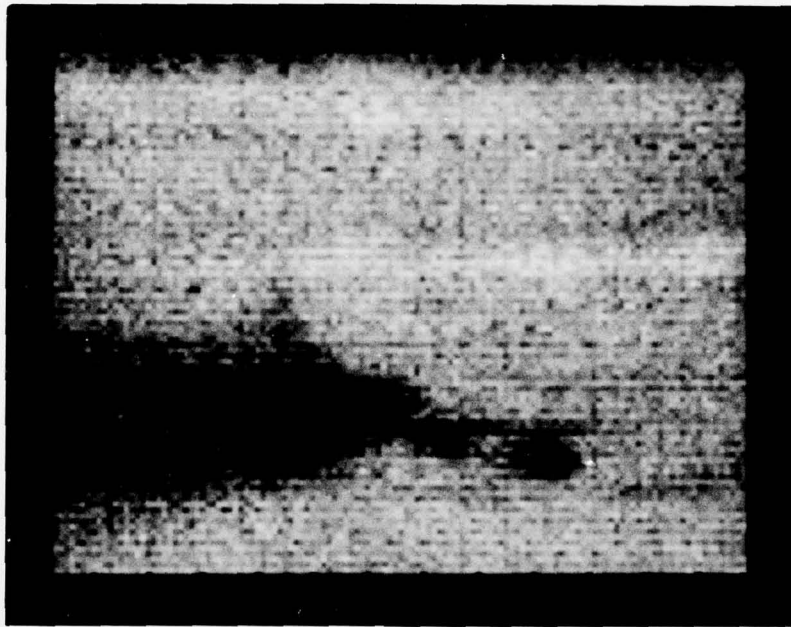
Frame 21
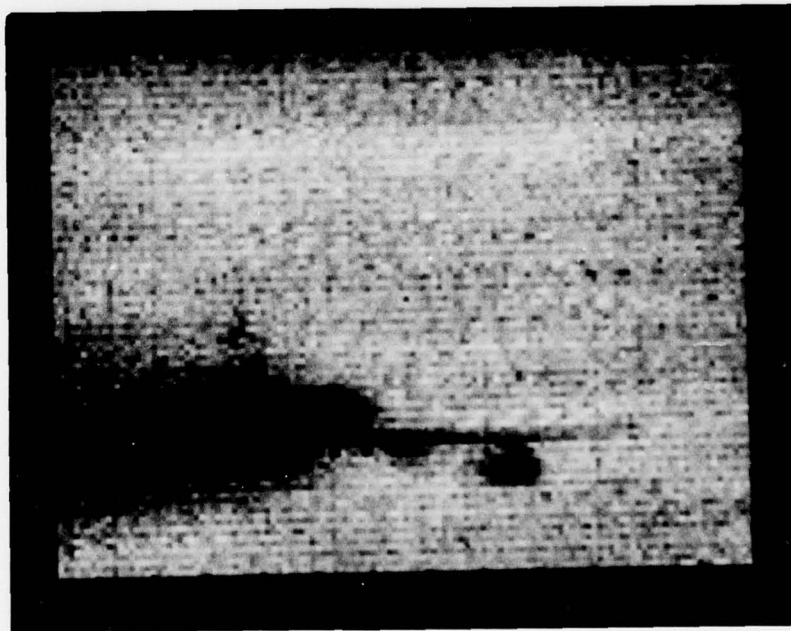
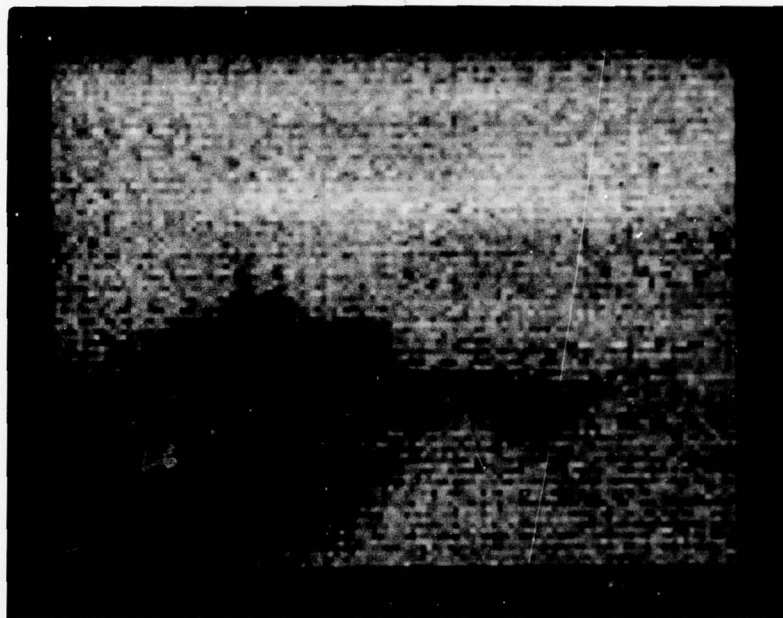
Frame 31

Figure 59. AIRPLANE - frames 21 and 31
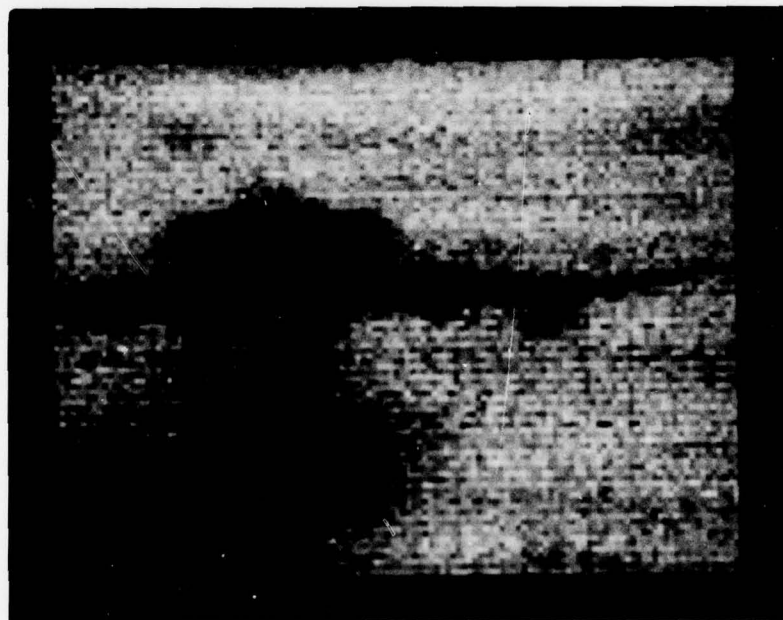
Frame 41

Frame 51

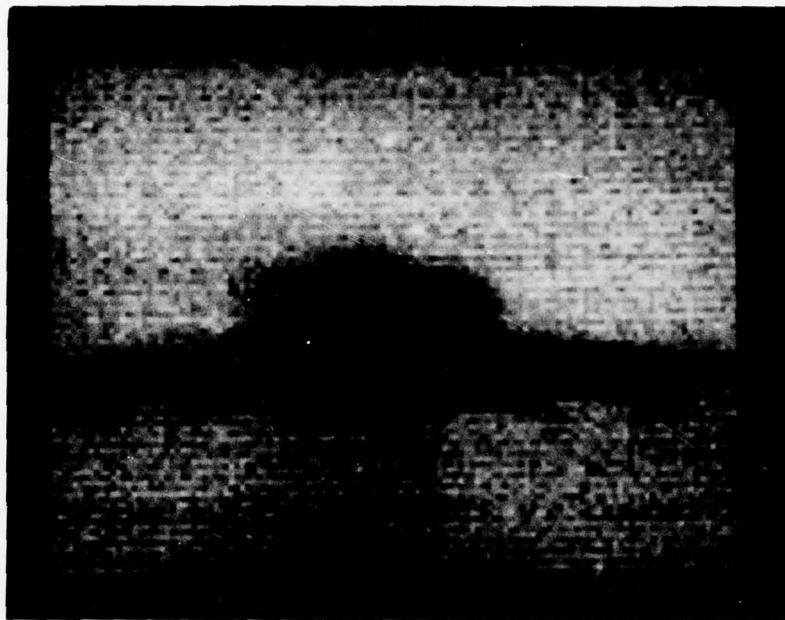Figure 60. AIRPLANE - frames 41 and 51

196

Frame 61

Frame 71

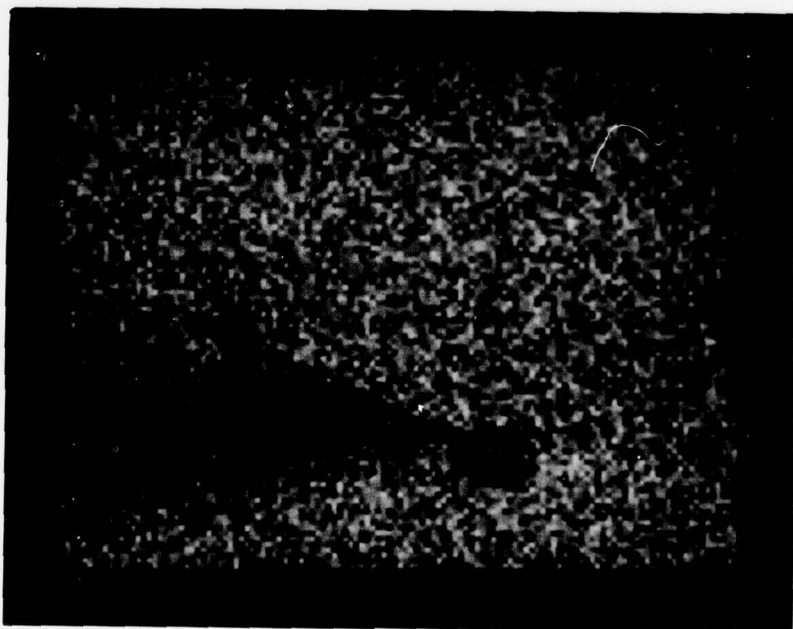Figure 61. AIRPLANE - frames 61 and 71

197

Frame 81

Frame 89

Figure 62. AIRPLANE - frames 81 and 89

198

Without noise

With noise, $\sigma_m^2 = 25$

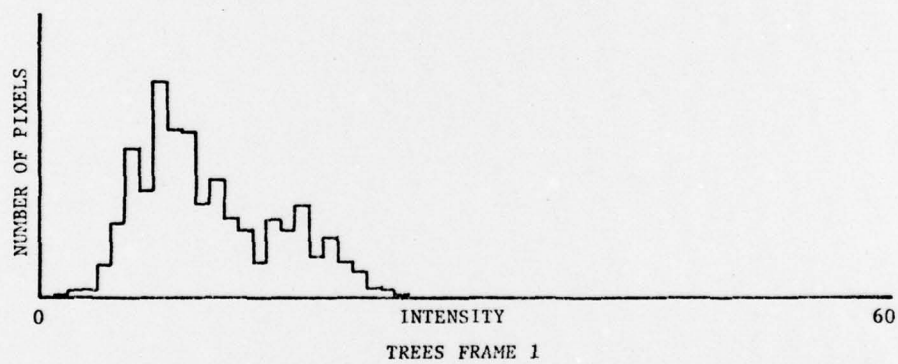Figure 63. Underlying image used for simulations with and without noise

199

NUMBER OF PIXELS

0   INTENSITY   60

CARS FRAME 1

NUMBER OF PIXELS

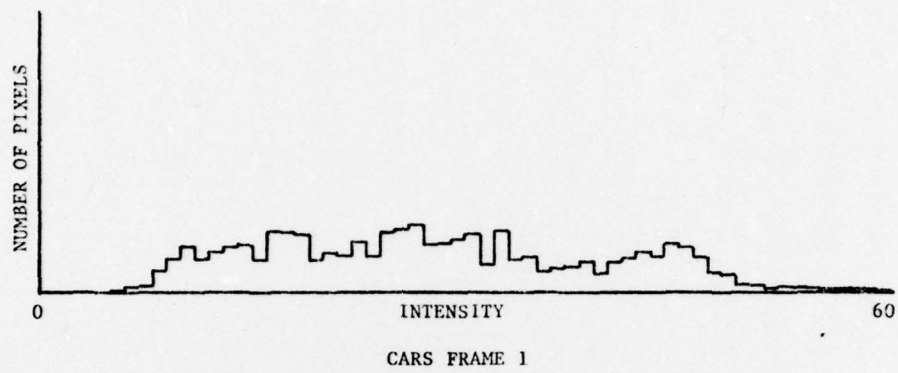0   INTENSITY   60

TREES FRAME 1

Figure 64.   Intensity Histograms for Frame 1
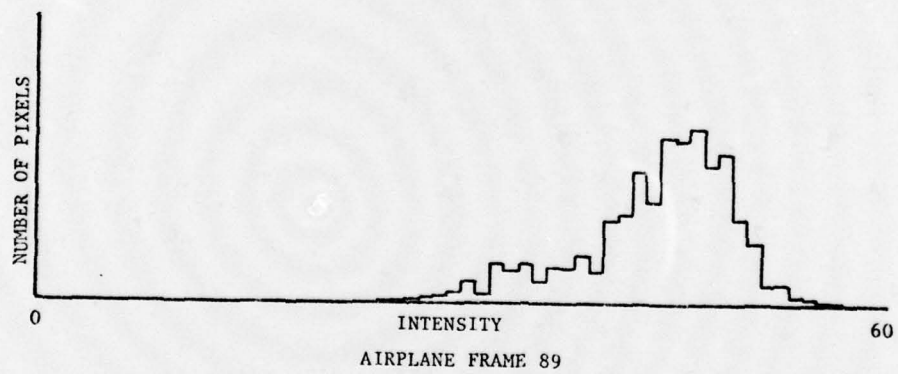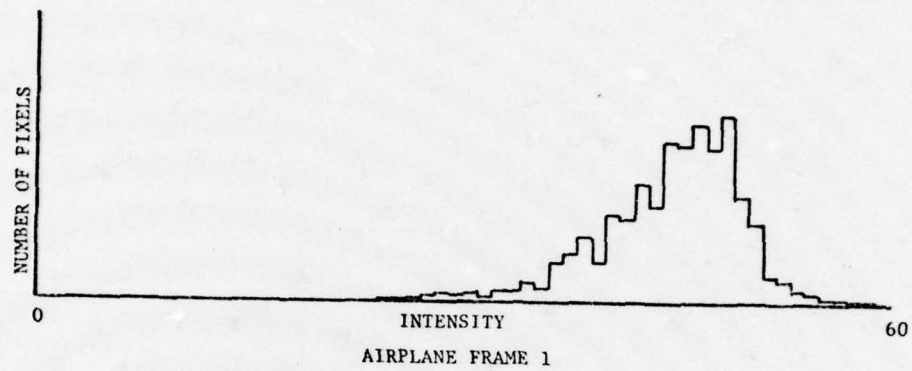of CARS and TREES

200

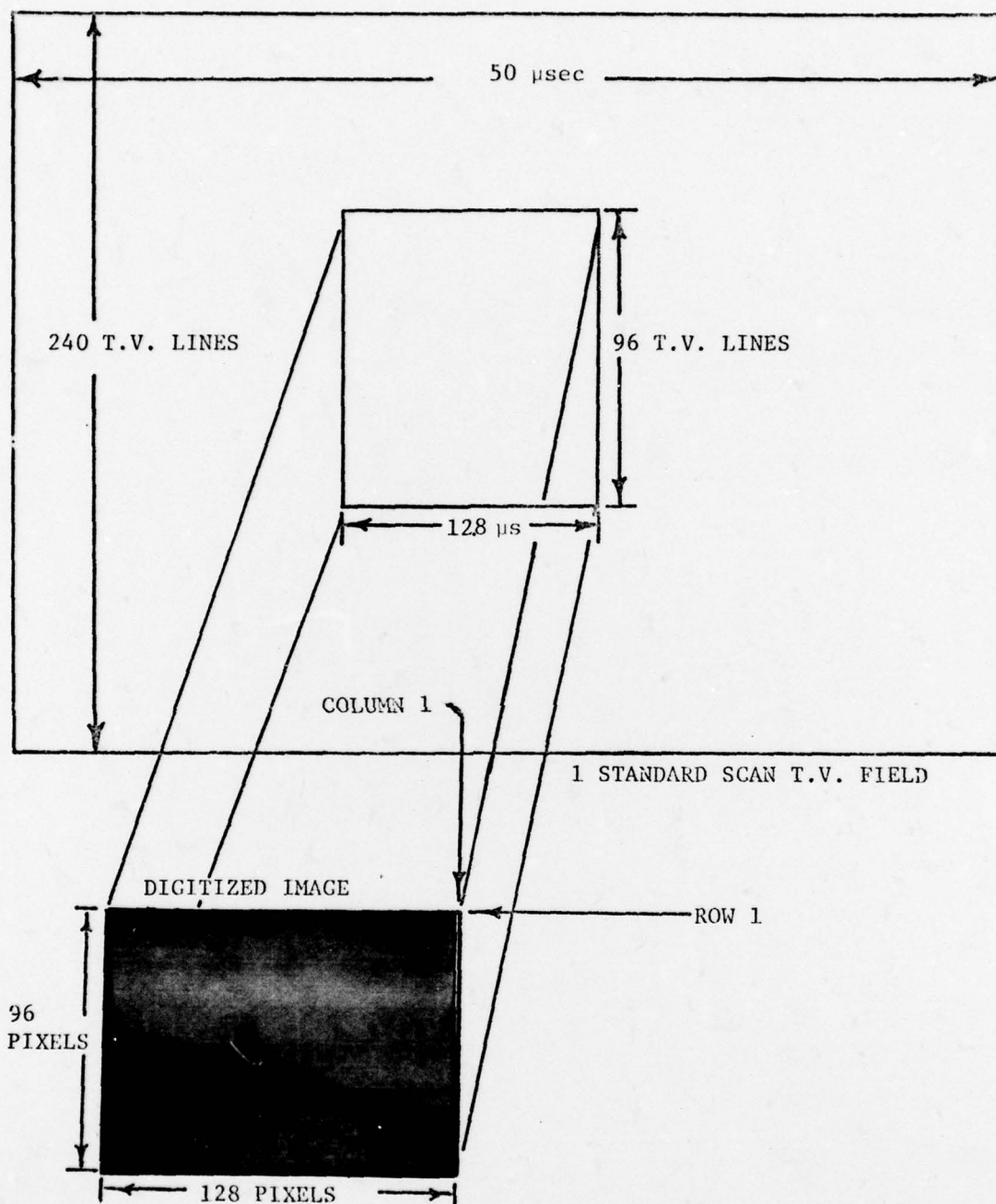Figure 65.   Intensity Histograms for Frames 1
and 89 of AIRPLANE

Figure 66. Relationship Between Standard
Scan T.V. Field and Digitized Images

VITA

Llewellyn S. Dougherty was born on 31 August 1944 in Portland, Oregon. He Graduated from Woodside High School in Redwood City, California in 1962 and entered the University of California at Berkeley in the fall of that year. In 1963 he was selected to attend the United States Air Force Academy, graduating in 1967 with a Bachelor of Science degree in Engineering Science and a commision in the Air Force. He attended the Massachusetts Institute of Technology on a National Science Foundation Fellowship, graduating in 1968 with a Master of Science in Aeronautics and Astronautics. He was assigned to the Guidance Test Division of the Air Force Missile Development Center, Holloman Air Force Base, New Mexico from 1968 until 1973. During this period he performed laboratory and flight tests on a wide variety of visible and infrared sensing terminal guidance tracking systems. In 1973 he entered the Air Force Institute of Technology, School of Engineering.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFIT/DS/EE/78-4	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) New Techniques for Tracking Sequences of Digitized Images		5. TYPE OF REPORT & PERIOD COVERED PhD Dissertation 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Llewellyn S. Dougherty, Captain USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology (AFIT/EN) Wright-Patterson Air Force Base, OHIO 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS same as 9 above		12. REPORT DATE October 1978 13. NUMBER OF PAGES 215
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE AFR 190-17.

JOSEPH P. HIPPS, Major, USAF
Director of Information   FEB 8 1977

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Correlation Tracker, Trackers, Two-dimensional Filters, Electrooptic Tracking, Picture Processing, Edge Detection, Image Registration, Kalman Filter, Similarity Detection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A model for a generalized image tracking system is presented. Characteristics of minimum norm similarity detectors are investigated. A first-order local tangent plane model for digitized imagery is used to successfully predict properties of the auto and cross distance functions for real data. A matrix signal-to-noise ratio is shown to be the natural signal-to-noise ratio for the minimum norm detection problem, and an approximation is derived and experimentally verified for an upper bound on the probability that a minimum

DD FORM 1 JAN 73 1473   EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

block 20 continued

norm detector makes a particular error. A non-linear two-dimensional
filter is presented which shows a significant reduction in noise variance
in low contrast regions of an image. An optimum weighted norm is derived
which minimizes the probability of making a registration error, and an
adaptive reference set selection algorithm is presented which maximizes
the tracking signal-to-noise ratio. The adaptive reference set selection
algorithm uses the histogram of gradient magnitudes and includes a new grad-
ient estimator/classifier with a fixed probability of error. An adaptive
Kalman filter is developed to update the reference image and the filter
is shown to be stable in all areas of interest.